Psychological Bulletin

Volume 51, 1954

CONTENTS OF VOLUME 51

ARTICLES

BOOK REVIEWS

MISCELLANEOUS

Psychological Bulletin

PSYCHOLOGICAL CONSEQUENCES OF BRAIN LESIONS AND ABLATIONS

SEYMOUR G. KLEBANOFF,[1] JEROME L. SINGER, AND HAROLD WILENSKY

Franklin D. Roosevelt Veterans Administration Hospital, Montrose, New York[2]

In a previous paper, the senior author (164) reviewed the literature dealing with the psychological consequences of organic brain lesions and ablations through the year 1941. The intervening period of over ten years has witnessed considerable research in this area. In addition to the need to scrutinize and systematize the new and extensive literature relative to the problem of therapeutic psychosurgery, it is also important to observe significant trends in the development of new psychological evaluative techniques. Also, there have been notable modifications in the use and application of existing test methods in organic brain disease.

In the previous review of the literature by Klebanoff (164), effort was made to formulate certain basic generalizations. It appears appropriate now to examine critically these

[1] Assistant Professor of Psychology, Department of Psychiatry, Cornell University Medical College.

conclusions in the light of the extensive research of the last decade. One is impressed by the striking development of new specialized test techniques. It seems important to evaluate these new techniques in terms of their potential clinical value. Research of the last decade has stressed the potential importance of such variables as the patient's age and the quality of interaction with the environment in producing differences in deficit in patients with apparently similar disease processes.

In the light of all the above considerations, the present review has been undertaken. The general attitude underlying this survey is one of critical appraisal rather than mere review; the basic objectives are to organize, integrate, clarify, and extract significant generalizations. Such an approach appears to be indicated if the basic conclusions in the area of psychological testing in organic brain disease and psychosurgery are to be evaluated in terms of essential contribution to scientific and clinical knowledge.

Brain Injury in Children

Psychological studies of persons who have suffered congenital brain injury, encephalitis, or other types of cerebral damage early in life have proven an increasingly fertile area of

research. The significance of research with brain-injured children, in addition to its practical value for rehabilitation, lies in the possibilities of obtaining valuable understanding of the developmental aspects of brain function or dysfunction. Theories of normal mental development (135, 180, 286) have stressed the fact that psychological functioning and cerebral organization show an increase in differentiation and integration with age through adolescence. Although the less differentiated structure in the child leads to a more diffuse effect of injury if it occurs early, it also makes possible maturation with compensatory development of function. This negates the influence of an injury which would frequently have serious long-term sequelae in adults.

Probst (222) followed up a group of 95 children with skull fractures occurring between ages one and fifteen. After seven to fourteen years there was no evidence of impairment in school or vocational achievement, social adjustment, or intellectual development. Several investigators (50, 64, 84, 126, 163, 297) obtained varied results in brief case studies of functions disturbed by brain injury in rather young children. French (84) studied ten children who were operated upon for cerebellar tumors with an extensive battery of generalized intelligence and specialized tests and found no evidence of impairment on deterioration-sensitive tests. Some children showed superior performance. On the other hand, Cotton (64), employed a control group in studying the thought processes of spastic children and found striking differences between them and the normal control group. It is possible that differences in severity of injury may well account for these divergent findings. An unusual case study described by Klapper and Werner (163) compared three children with cerebral palsy with their normal identical twins, using a large battery of tests sampling a wide variety of functions. The brain-injured children were impaired in every area when compared with their normal twins.

The major portion of research on brain injury in children has emerged in the extensive work on the distinction between endogenous and exogenous feeblemindedness. This distinction has proven of great heuristic value in stimulating an intensive study of the various dysfunctions consequent upon brain injury. As a result of this research, it has been possible to develop fairly adequate criteria for distinguishing these two types of defective children (32, 176, 247, 256). This has resulted in the addition of a number of fairly sensitive and important techniques to the repertory of the clinician for measuring basic functions likely to be impaired in early brain injury.

Motor Functions

The brain-injured child is generally subject to motor disturbances, since either cerebellar injuries or disturbances in the motor cortex are common (32, 176, 247). Even when obvious motor disturbance is not present, the brain-injured child shows awkwardness and incoordination in performance. Heath (130, 131), comparing exogenous and endogenous feebleminded children of comparable mental age in their performance on a simple rail-walking test, found that the exogenous children proved strikingly inferior. Descriptive studies of the motor performance of these exogenous feebleminded characterize them as hyperkinetic, overactive, restless, impulsive, and inco-

ordinated (247, 283). The impression gained from such descriptive studies suggests that motor difficulties are as much a function of fairly subtle disturbances in the integration of behavior and the inhibition of overt responsiveness as they are of specific damage to motor areas of the cortex. The use of the Van Der Lugt Scale (270) seems to offer potential for major research in the area of motor development and defect in the future.

Perceptual Functions

The most extensive and definitive studies of brain-injured children have been in the area of perception. This work has been of theoretical as well as of practical importance since it has demonstrated striking relations between perceptual and motor functioning that has led to suggestions for a sensory-tonic theory of perception (248, 282). Thus, Strauss (255), in summarizing three case histories, described perceptual and thought processes which mirror the forced responsiveness, hyperkinesis, and incoordination of the brain-injured child in his motor performance. Bender and Silver (33) describe the case of a brain-injured child with a severe modification in perception of the body. They conclude that the body image of the brain-damaged child is disturbed by "tonus pulls," equilibrium problems, and perceptual and integrative difficulties which heighten his social inadequacy.

Werner and Strauss (292) studied normal children equated for mental age with endogenous and exogenous mentally defective children without gross motor or central visual defects. They used tachistoscopic presentation of marred figures and copying of patterns of marble boards or reproduction of the patterns by drawing. Although the endogenous children

were generally inferior to the normal group. They differed strikingly in mode of response from the exogenous children. The latter made a great many errors that suggested failure to discriminate between figure and ground.

Subsequent reinforcement of these findings has come from the investigations of Lord and Wood (186) and Dolphin and Cruickshank (70). The latter studied perception of embedded figures and mosaic reproduction by matched normal and cerebral-palsied children of low-average intelligence. They attributed the significant inferiority in perceptual performance of the children with cerebral palsy to forced responsiveness and meticulosity. Bensberg (42), in a careful study, compared familial mental defectives and brain-injured defectives using the marble-board procedure. The brain-injured group proved significantly inferior in accuracy and tended more often to show "jumps" to new lines in their drawings. The results continue to point to pathological figure-ground perception as a basic difficulty consequent on brain injury. Werner (283), studying the performance of the two feebleminded types on the Rorschach, noted significantly more white space responses (S) for the exogenous group. Case studies by Strauss and Werner (257, 258), and Werner and Carrison (289), demonstrate ingeniously the interrelationships of defective figure-ground formations in brain-injured children and other symptoms such as finger agnosia and arithmetic disability.

Similarly, these brain-injured children showed inability to grasp and retain visual patterns made up of discrete elements. The authors attributed this to a basic difficulty in integrating several aspects of a visual

stimulus. This is a defect similar to that reported for a brain-injured adult by Nichols and Hunt (205) and considered a basic consequence of severe brain pathology by Hunt and Cofer (145) in their review of psychological deficit. A case study by Schilder (243) has similar possibilities for interpretation of this deficit.

Brief mention may be made of additional perceptual difficulties manifested by brain-injured children. Werner and Thuma (295) compared equated groups of feebleminded representing the two etiological types on performance in critical flicker fusion perception. Results showed the exogenous feebleminded fell significantly below the familial group in critical flicker frequency at all three brightness levels. This finding is quite important in view of extensive research on adults reported by Teuber (263) and Halstead (120). Similarly, a marked defect in the perception of apparent or illusory motion, the phi phenomenon, and tachistoscopic presentation of stick figures in imbalance was found among the exogenous feebleminded by Werner and Thuma (294). These children also showed fewer movement percepts on the Rorschach than an equated group of endogenous feebleminded. In one of the few studies involving auditory perception in brain pathology, Werner and Bowers (288) found that the brain-injured feebleminded were deficient in discriminating melodic patterns, revealing much the same disturbance in figure-ground pathology noted above.

Thought Processes

Sorting tests have proven increasingly useful in studying thought in process and in its effect. Strauss and Werner (259) employed the Halstead sorting test to determine if the findings with brain-injured adults were comparable to those obtained with brain-injured children. Twenty pairs of exogenous and endogenous feebleminded children were matched for mental age and intelligence. The brain-injured children formed more relationships between objects in the voluntary sorting but accomplished this by forming "singular and unusual combinations." In choosing objects to go with rather unambiguous pictures, the brain-injured children also selected uncommon objects. They deviated more from standard names, organized objects into circumscribed units, overstressed orderliness, and showed extreme concreteness. In contrast with the endogenous children and particularly with a group of healthy children, the brain-injured children were attracted to "properties of objects apt to elicit motor responses." The latter finding seems important in view of the hyperkinetic tendencies ascribed to brain-injured children.

In a further study of thought processes, Werner and Carrison (290) employed a standard procedure for studying animism or the tendency to perceive natural events or objects as living. The results, based upon levels of thought development derived from the work of Piaget, showed significantly more animistic responses by the exogenous children. In keeping with certain concepts of Goldstein (93) and Werner (286), the animistic tendencies of brain-injured children may be looked upon as functions of a greater rigidity and concreteness which prevent them from detaching themselves from objects and events. They are thus unable to differentiate between their own feelings and those of the surrounding world. Evidence that these tendencies are a result of impairment due to brain injury

rather than merely representing a particular type of intellectual limitation is forthcoming in a study by Cotton (64), who compared physically normal children with spastic children of average intelligence. Results on a series of sorting, verbal completion, and patterning tests showed a decided inferiority for the spastics whose responses were more stereotyped and who were less inclusive in sortings and more given to highly personal associations and other signs of concreteness. These findings were particularly associated with speech difficulties and were found less often when the congenital injury was restricted to motor areas.

The problem of the rigidity or perseveration of brain-injured or feebleminded generally has been discussed theoretically by Lewin (180) and Kounin (173). The latter carried out several significant studies demonstrating differences in rigidity patterns as a function of chronological and mental age. It is not clear whether the subjects used by Kounin were exogenous or endogenous feebleminded, although apparently they were physically normal as determined by medical examination. Werner (284), summarizing four experiments comparing rigidity patterns of brain-injured and familial feebleminded, observed more perseveration in the former group. He characterized the rigidity of the brain-injured as *abnormal;* it involved a trend away from the global approach toward isolation of elements of a series which became self-contained and were repeated monotonously or "jumped" suddenly into the foreground. The familial feebleminded, on the other hand, showed *subnormal* rigidity related to a relative lack of differentiation. They showed predominantly global behavior, perception, or action, organized as undifferentiated wholes so that situations not sharply set apart were easily fused, resulting in stereotyped responses.

Personality

There have been comparatively few attempts at extensive personality evaluations of brain-injured children. Werner (283) studied the Rorschach test performance of brain-injured feebleminded and found that they showed a restriction of creative fantasy activity, strong tendencies toward explosive emotionality, negativism, and the disintegrative behavior described above. Colm (62), employing the Lowenfeld Mosaic Test, compared normal and brain-injured children. She described the brain-injured children as "stimulus bound," with an inability to shift and a repetitiveness which indicated an impairment in abstraction. Thus, these children showed simple additive placement of the mosaic pieces using side-by-side or color-by-color patterns and showed difficulty in formulating and executing a plan for a design. Colm indicated an important distinction between the exogenous children and compulsive or autistic children who also showed repetitiousness. The latter perseverated chiefly on words with strong emotional content while the brain-injured children repeated almost any phrase regardless of the emotional tone.

In conclusion, the extensive research on brain-injured children indicates defects that are generally like those described for brain-injured adults. The children show inability to develop abstract functions, pathological rigidity, lack of creative imagination, stimulus bondage, and disintegrative behavior in all areas. Brain-injured children also show strikingly

unequal development of capacities, suggesting some cerebral localization of damage, but the basic defects in abstraction hinder any possibility of really integrated development for many.

ORGANIC PSYCHOSES AND DE-GENERATIVE DISEASES OF THE NERVOUS SYSTEM

Despite the increased interest in problems of the aged in the past decade, there has been little extensive research dealing with psychological deficit or emotional changes concomitant with senescence. A major trend in the past decade has been an awareness of some decline in intellectual functioning with increased age (151, 274). There has been little attempt, however, at precise analysis of the nature of this decline and its relationship to emotional stability and the psychological milieu.

Fortunately, some studies on the organization of mental life in organic psychoses are beginning to appear. Botwinick and Birren (49), for example, compared the intellectual performance of hospitalized seniles with that of normal persons of the same age range (60 to 70 years) and with that of a matched group of young normals. They used the Wechsler-Bellevue and Babcock-Levy Revision to test for deterioration. These investigators concluded that while differences between the two aged groups existed, the older persons were far more like each other than were the normal aged and the younger group.

Of greater theoretical interest, perhaps, are studies attempting to delineate the pattern of mental change consequent on the fairly diffuse degenerative brain changes that seem to occur with cerebral arteriosclerosis and similar diseases. Thus, in studies comparing 100 male senile dements

with an average age of 73, and employing such tests as Raven's Progressive Matrices, Eysenck (72, 73) found considerable deterioration in the patients in terms of healthy adult and child norms. A factor analysis of the results showed a general factor and three factors relating to speed, memory, and physical strength which could be identified fairly readily. Tests of abstract reasoning proved most sensitive to deterioration in contrast to tests involving memory and stable knowledge. A differential deterioration in abilities appeared to occur along the lines suggested many years ago by Hughlings Jackson (149) in which the diffuse cerebral damage leads to initial loss of the most complex functions which develop most recently in the evolution of the nervous system. In this connection, Halstead (115) studied senile dements over 70 with an extensive battery of psychological tests. There were two groups in different stages of dementia. Tests discriminating between the two groups involved the more complex functions. The tests included Porteus Mazes, Knox Cubes, Block Designs, and various recent memory tests.

The impoverishment of functioning in senile psychotics was brought out in a study by Cleveland and Dysinger (57). These investigators studied the performance of institutionalized seniles, whose average age was 75, with the Wechsler-Bellevue scale and an object-sorting test. The patients manifested decided difficulties in assuming an abstract attitude and could not sort on a conceptual basis. That even verbal behavior suffers impairment in senile dementia was demonstrated by Feldman and Cameron (75) and Ackelsberg (3). The former investigators compared the speech of senile psy-

chotics with that of normal adults and children. Although based on rather small samples, this study does suggest that seniles have difficulty in dealing with abstract terms and rely most heavily on concrete nominal forms or words of action to the neglect of adjectives which involve imaginative processes. Ackelsberg (3) obtained comparable findings in vocabulary studies of seniles. Recent studies by Pinkerton and Kelly (217) and Hall (113) contrasted deteriorated seniles with children and depressives by the use of various sorting tests and supported the thesis that a loss of conceptual ability occurs early in the course of brain degeneration while more general intellectual test performances remain relatively intact.

Diffuse degeneration of the cortex, particularly in the frontal areas, occurs in general paresis. The behavior of paretics in the classic description of the disease, while more flagrantly psychotic, does resemble that of the older senile dementia or arteriosclerotic patients. With the improved methods for treating syphilis and arresting general paresis, a re-evaluation of the description of the disease may well be necessary.

Psychological studies of patients with general paresis have been fairly frequent in the past decade and have provided suggestive findings. Studies with the Wechsler-Bellevue scale (88, 195) as well as with the CVS abridgement (198) have suggested the usefulness of these tests in delineating the intellectual performance of paretics. While there were individual differences in the degree of impairment on the various subtests, the general findings suggested lowered concentration ability, greater difficulty in new learning, verbal comprehension, and concept formation in the case of the paretic group. Trist, Trist, and Brody (269) used a large battery of cognitive tests in comparing normals, neurotics, and paretics and observed that the paretics found it difficult to ignore irrelevant detail, and a flowing together of figure and ground was evident in their performance.

A somewhat novel approach to the psychological study of paresis is offered in studies by Rashkis, Cushman, and Landis (225) and Rashkis (224). The former study involved the administration of a word-sorting test to normal adults, normal children, general paretics, and schizophrenics equated for educational achievement. On the basis of an analysis of the sortings, only the normal adults were able to assume the abstract attitude. Children and schizophrenics functioned at the complex, "pseudoconceptual" level, while the paretics were unable to reach even that degree of attainment. From the standpoint of volitional attitude, the schizophrenics more closely resembled the normal adults while the paretics resembled the children. Rashkis (224), employing more careful controls, amplified and extended these findings. Rashkis studied three groups consisting of schizophrenics, general paretics, and cerebral arteriosclerotics. The schizophrenics proved somewhat better "coordinated" and more capable of accounting for their performance. The arteriosclerotics, although "uncoordinated," also attempted to account for their performance, in contrast to the paretics who were "uncoordinated" and offered no excuse or apology.

In one of the few attempts to study the personality characteristics of general paretics, Klebanoff (165) employed the Rorschach inkblots to compare findings between a homo-

geneous group of paretics and a normal control group matched for age and education. The paretics lacked sufficient intellectual drive to function at an abstract level. They showed concrete mental activity which lacked accuracy, conformity, and "fundamental adaptivity to environmental needs."

A promising and somewhat different approach to this problem is suggested by the preliminary study of Wittenborn, Bell, and Lesser (300) who carried out a factor analysis of symptom ratings along various scales for deteriorated organics, hebephrenics, and young patients with functional psychoses. Different symptom clusters emerged with only anxiety and paranoid factors coinciding for the groups. Deterioration was most marked among the organic patients.

Research is just now beginning to appear in connection with presumably subcortical neurological degeneration such as is found in multiple sclerosis. Although somewhat contradictory results have emerged with respect to possible mental deterioration concomitant with multiple sclerosis (25, 53, 68), differences in the populations studied and in the duration of the disease may account for these discrepancies. Canter (53), for example, found definite evidence of deterioration among patients in the early phases of the disease when compared with both their own military induction test scores and a control group of clinic employees. Diers and Brown (68) reported no evidence of decline, while Baldwin (25) reported loss of abstract ability in some patients and no apparent decline for many others.

It is in the area of personality that a striking uniformity emerges in three separate studies. A brief report by Harrower (127) was based on comparisons of Rorschach and Szondi findings for multiple sclerotics, control groups of normals, emotionally unstable psychosomatic patients, and patients with poliomyelitis and Parkinson's disease. It was found that the multiple sclerotic patients differed from the other groups in showing *least* concern with bodily symptoms, extreme dependency, resignation, denial of conflicts, and a need to excite sympathy. This general similarity to the classical picture of the hysteric patient with *belle indifference* also emerged in Rorschach studies by Blatt and Hecht (48) and an intensive study by Baldwin (25) of family background, premorbid factors, and data from Minnesota Multiphasic Personality Inventory patterns. The general similarity of the multiple sclerotic and neurotic hysteric is particularly striking in the early phases of the disease (25) and suggests the importance of a premorbid personality factor. In view of similar findings with idiopathic epileptics (91), an important area of research in the psychosomatic implications of neurological disease appears to require extensive exploration.

BRAIN TUMOR, SURGICAL ABLATION, AND CEREBRAL TRAUMA IN ADULTS

Disturbances in Perception

The majority of studies have been restricted to the field of visual perception. For excellent reviews of material on the borderline of visual sensation and perception, the reader is referred to Bender and Teuber (40) and Teuber and Bender (267). Halstead (120) has ascribed important theoretical significance to the critical flicker fusion frequency as a result of factor analyses of the test performances of normal individuals. On

the basis of intercorrelations between the critical flicker frequency and performance and other techniques, Halstead (120) has tentatively defined a power factor (P) which indicates the kinetic capacity of the individual, not unlike the "central vigilance" described by Head (128) in his studies of aphasia. Halstead (117, 119) has reported evidence that individuals of equal IQ may differ markedly in the P factor, which is sensitive to disruption by the presence of relatively small lesions in the brain and under conditions of low-grade anoxia or fatigue. Since Halstead indicates that a lowered critical flicker frequency is one of the prime indicators of a depressed level of the P factor of biological intelligence, the value of this simple laboratory procedure for psychological evaluation of the consequences of brain lesions is greatly enhanced.

Unfortunately, the findings of Halstead have not been fully confirmed. Battersby (27, 28) and Battersby, Bender, and Teuber (29) obtained negative results in studies of veteran patients with frontal lobe head wounds. These patients failed to show any significant differences in critical flicker frequency level when compared with an equated control group. On the other hand, a series of studies (28, 29, 38, 40, 265, 266, 303) has shown definite lowering in critical flicker frequency after occipital lobe lesions. This impairment was detected in areas that appeared normal under routine perimetric examination. Bender and Teuber (38, 40) have reported that defects emerged more strikingly when critical flicker frequency was tested in the periphery of vision rather than in the macular region. These results, while confirming those of Halstead (120) in raising doubts concerning the significance of

an anatomic point-to-point projection in visual structures, are difficult to reconcile with his observations on frontal lobe patients. The differences may be a function of procedural variations or, as Battersby (28) has suggested, the result of differing patient groups.

Case studies comprise the bulk of the studies on perceptual response following brain injury. Rather challenging for general theories of perception are observations of a phenomenon tentatively labelled "extinction," in which patients confronted with an object in their left-half field report that it disappears as soon as another object is shown in their right-half field (34, 36, 303). As Hebb (136) has noted, these phenomena are difficult to assimilate into current perceptual theories. Bender and Furlow (34) describe the manner in which tendencies to reorganize a psychological field of vision about a subjective center make it difficult for a patient with occipital lobe damage to recognize that he has lost central vision.

Problems of unusual spatial reorganizations brought out only by special testing devices are described in a number of cases of occipital injury. Paterson and Zangwill (211) observed in two cases of right parieto-occipital lobe injuries a tendency to overestimate the distance of very near objects and to underestimate the distances of far objects, although the ability to appreciate depth and distance and the "implicit awareness of space" were not affected. Analysis of spatial structure was difficult, and various spatial relations tests and the Kohs Blocks brought out the defects in the patient. Drawings of complex objects were carried out piecemeal with poor articulation of subwholes. Temporal disorientation, also found by Coheen (58) with a larger number

of patients and various control groups, was also present. Since most case studies in this group represent traumatic injuries, a report by Stengel (254) of consequences of vascular lesions with an eclampsia in a 40-year-old woman is interesting since quite severe spatial reorganization occurred. Extensive testing revealed that for this subject the complex spatial organization of the outside world had been replaced by a most primitive organization in which "nearness" served as the only measuring scale.

A series of cases of parietal and occipital lobe injuries reported by Bender and Teuber (38, 40) is particularly valuable because of the careful attempts by the authors to relate findings to a general theory of brain function. On the basis of an intensive sampling of increasingly complex perceptual performances, the authors report such defects as disturbance (limited to the lower left quadrant for some patients) in localizing objects, teleopsia with a tendency toward an excess of depth in the subjective field, higher threshold for phi, lowered critical flicker frequency, and tendency to enlarge unfamiliar objects. For some patients the subjective coronal plane was rotated as a whole, away from them on the left, toward them on the right. This shifting could be compensated for by one patient in daily life, while for another, subjective visual space no longer coincided with his tactile or locomotor space and led him into many constant errors in pointing. The authors feel that these results argue against associational theories and in favor of "vector" theories generally related to Köhler's (170) views of brain function.

In an experimental attempt to test certain hypotheses concerning cerebral function related to Köhler's theories, Klein and Krech (166) and Jaffe (150) independently carried out recent studies of the kinesthetic figural aftereffects in matched brain-injured and control groups. Jaffe, using patients with traumatic head wounds, found no differences between controls and patients. Klein and Krech tested the hypothesis that both concreteness and figural aftereffects in the brain-injured were functions of reduced "cortical conductivity" (174). They predicted more pronounced figural aftereffects that would appear sooner and last longer for subjects of varied degrees of brain pathology than for equated controls. Results supported the hypotheses and, in addition, high positive correlations were found between neurologists' ratings of extent of damage and the vividness and duration of the effect. The differences in results of the two studies may be in part a function of different patient populations, traumatic head wounds as against surgical ablations, a persisting obstacle to the reconciliation of data in this field.

There has as yet been little effort to relate perceptual defects in brain-injured patients to more general areas of functioning such as motivation and personality. Critchley (66) has stressed the relation of body image to alteration in cerebral functioning following parietal lesions. Halstead (118) has shown a possible relationship of lowered and fairly rigid critical flicker frequency and poor performance on the Dynamic Visual Field test to severe judgmental defects in a man following unilateral lobectomy. More recently, Weinstein and Kahn (278, 279, 280) have investigated motivational and interpersonal aspects of the perceptual response of 50 patients with bilateral lesions of neoplastic or vas-

cular origin. These patients showed distortions not only in body image, but in self-concepts and manifested tendencies to deny illness or impairment and to misperceive stimuli related to their disability or hospitalization. The denial of affective involvement sometimes reported following lobotomy may represent a similar phenomenon and quite a serious threat to future adjustment since it substitutes an equally unrealistic self-orientation for the previous "tortured self-concern."

Summarizing results in psychological studies of perceptual functioning following brain lesions and ablations, it is apparent that the movement toward specialized testing techniques pointed out by Klebanoff (164) has progressed rapidly. Perceptual research with brain-injured patients suggests that the repertory of the clinical psychologist in the examination of neurological patients may soon be augmented by the use of specialized laboratory methods such as tachistoscopic presentation, critical flicker fusion frequency testing, apparent movement thresholds, and other so-called "brass instrument" techniques. The value of these perceptual techniques in localization of lesions remains equivocal. It seems clear, nevertheless, that subtle perceptual impairments or reorganizations following cerebral injury are most clearly brought out by intensive laboratory examination rather than by the widely used global clinical methods like the Rorschach, Bender-Gestalt, Wechsler-Bellevue, Hunt-Minnesota, etc. These laboratory techniques or those outlined by Goldstein (95) are now chiefly available in only a few research centers (120, 263, 307). Before long, with the aid of these new techniques, the clinical psychologist may be in a position to play an ever more vital role in a neurological setting.

Memory and Attention

The past decade has provided comparatively little intensive research on specific memory or attention disturbances in brain-injured patients. The emphasis, rather, has been on incorporation of these functions within broader categories such as perception or general intelligence. Certainly, many disturbances of immediate memory may actually reflect difficulties in attention to presented material. The distractibility and stimulus bondage which are reported to be characteristic of both brain-injured adults and children (93, 94, 291, 293) would naturally impede original learning of material. Hall and Crookes (114) compared a heterogeneous brain-injured group with normals and schizophrenics in learning ability and found some qualitative suggestion of impairment in the organic patients, although differences between the organic patients and the schizophrenic patients were not definitive. Ruesch and Moore (234) studied 190 patients immediately following head injury and reported that the serial subtraction test ("100–7") proved most sensitive, suggesting that the ability to maintain sustained effort is particularly sensitive to cerebral dysfunction, whether momentary or persistent. Similarly, other investigators have found evidence of memory loss dependent largely upon failures in attention and immediate retentiveness (4, 10, 20). Tests such as Halstead's Formboard Retention and Dynamic Visual Field (120), Teuber's Field of Search (263), and Benton's Visual Retention (43) have revealed fairly clear evidence of attention and immediate memory defects in cases of brain trauma or

surgical ablation. Definitive standardization of these procedures in the future should prove them valuable clinical aids. The major problem of devising laboratory and clinical procedures capable of discriminating between certain schizophrenic patients and individuals with various kinds of brain pathology remains a difficult methodological consideration.

Mention should be made of some tests designed to detect brain damage by memory techniques (69, 102, 144). Hunt (144) has developed an extensive battery employing as a base the 1937 Stanford-Binet with verbal and nonverbal learning and recall tests. The test was developed on patients with diffuse brain injury, but subsequent research (4, 20, 154, 196) suggests that the test does not discriminate well.

Employing an omnibus memory test approach, Graham and Kendall (102), while reporting significant differences between a brain-damaged group and a neurotic group, found impairment present in only half of the organics. In a careful study by Cohen (59) no measures of the Wechsler Memory scale discriminated between neurotics and patients with intracranial pathology. These negative findings do not necessarily indicate that memory defects are uncommon among patients or that memory scales are without value in neurological testing. These scales have undoubted usefulness as part of an individual testing battery in the hands of an experienced clinician. It is clear, however, that the complexity of brain functioning and the subtlety of psychological sequelae of organic brain impairment severely limit any research or clinical studies with omnibus memory scales.

General Intelligence

Although reports on general intellectual consequences of brain injuries or diseases are considerable in quantity, the past decade has seen little progress in the treatment of this issue. Critical case studies (129, 132, 202, 207) have suggested that general intelligence as measured by standard scales may be normal or actually improve following large but clean excisions of previously pathological brain tissue. Concern in employing intelligence tests has been, with certain exceptions, limited to the Wechsler-Bellevue Intelligence Scale and various measures of deficit derived from this test. Recognizing that intelligence in general may not show a change after brain lesion, investigators, following the clinical suggestions of Wechsler (275), have attempted to tease out patterns of test performance capable of differentiating the brain-injured or diseased from normal individuals. Unfortunately, as Rabin and Guertin (223) have noted in their review of Wechsler-Bellevue research, the Mental Deterioration Index (MDI), originally developed on the basis of decline in scores with increasing age among normals, has been translated too hastily to indicate decline following organic brain damage.

A brief report on clinical cases by Levi, Oppenheim, and Wechsler (178) indicated, for example, the practical usefulness of the MDI in differentiating patients with organic lesions from patients with hysterical and other psychogenic disturbances. Although cautious in their generalizations, these authors did raise an important principle for use in studying deterioration. The concept of tests that "hold up" with age and hence may conceivably be less resistant to

the inroads of brain damage is a challenging one. The question remains, however, as to the specific nature of "hold" and "don't hold" tests and to the quantitative relationships necessary to suggest pathological deterioration. In its purely quantitative form, the MDI has not been shown to be sufficiently discriminating in individual determinations of brain-injured and normals (8, 15, 47, 110, 156).

It should be noted, however, that the general principle of the contrast of "hold" and "don't hold" tests seems to be operative and that the majority of brain-injured are properly diagnosed by its use. For individual prediction, however, reliance on quantitative scores alone seems relatively futile. Anderson (15, 16) compared clinically substantiated cases of focal lesions in the dominant hemisphere with comparable cases of lesions in the nondominant hemisphere. He found that while the MDI did not yield very reliable quantitative discrimination, differences were in the expected direction. Brain-injured patients with focal lesions of the dominant hemisphere were more easily picked up by the use of the index than were those with nondominant lesions. On the basis of several studies and a comparison of brain-injured and brain-diseased patients, Allen (7, 8, 9, 10, 11) suggested a modification of the index, but other studies have provided data which suggest that this modification is not significantly more discriminating than the original (4, 47, 230). Other attempts at developing deterioration indices (138, 227) have been evaluated by Gutman (110), who found that a rather complex method developed by Hewson (138) agreed fairly well with clinical diagnoses of brain damage.

Far more significant perhaps than the empirical derivation of complex indices (which are not very discriminating and have not been cross-validated by the investigators) is evidence that certain functions appear again and again in the list as particularly susceptible to impairment by brain injury. Almost all investigators have noted that Digit Symbol, representing disciplined psychomotor learning, is most clearly affected. Also easily impaired by brain injury are Digit Span, a test of recent memory or attention (7, 9, 227, 233, 234); Block Design, testing analytic and synthetic capacities (4, 7, 9, 92, 107, 183); and Arithmetic (4), testing concentration and simple numerical facility. Since these subtests are most easily affected by anxiety or other emotional disturbances, a great deal of overlap between various clinical groups (223) is almost inevitable.

Too much of the research in this field has been excessively test-bound and empirical, relying heavily upon omnibus scales of general intelligence. New concepts like Cattell's "crystallized" and "fluid" abilities (54), similar views of Hebb (135) related to his physiological rationale, and Halstead's four factors as components of "biological intelligence" (120) should provide useful theoretical bases for more precise explorations.

Halstead provided some data to suggest a differential influence of brain lesions on his four factors of biological intelligence, with the P factor considered the most sensitive and disturbed, particularly in prefrontal lobe lesions, in his data. Halstead's results also suggest a gradient of impairment in biological intelligence with the frontal areas most sensitive. This hypothesis needs extensive testing since somewhat contradictory data are available from

other studies (29, 40). Nevertheless, Halstead's work poses a challenge to clinical psychologists to reconsider their techniques and their excessive reliance upon omnibus tests or techniques standardized for other purposes.

Disturbances in Thought and Language

In the past decade, research has continued to suggest that a major consequence of brain lesions and ablations is a disturbance in thought and language, classified as a loss of abstract or conceptual ability in the viewpoint popularized by Goldstein (94, 100). The directions taken by research in this period have involved (a) more precise exploration of this basic defect by the use of refined procedures (100, 113, 120, 121, 141, 264); and (b) attempts to ascertain whether the loss of abstract ability was particularly a consequence of frontal lobe pathology.

Extensively studied cases of both frontal and diffuse brain pathology have been reported (2, 44, 118, 123, 205, 306) in which patients manifested fairly general signs of impairment in abstract behavior as measured by procedures like the Weigl color-form sorting, the object-sorting tests, the Shipley-Hartford Conceptual Quotient, the Kohs Blocks, etc.

A striking and well-studied negative case was presented by Hebb (134) in which the patient was operated upon for removal of frontal scar tissue that had led to near-psychotic behavior. This patient improved subsequent to a fairly clean but extensive bilateral frontal lobectomy. He showed no clear signs of unusual disturbance in abstract behavior when presented with a host of psychological tests, many of which tapped conceptual processes. Differences between Hebb's case and cases discussed by Goldstein (98) may lie in the nature of the injury and the possible persistence of pathological tissue in some instances.

The value of various tests of abstraction in clinical neurological testing has been brought out in a number of studies (95, 97, 108, 139, 307). Greenblatt, Levine, and Atwell (108) found that, in comparing patients with known heterogeneous types of brain damage and patients without brain damage, abstraction tests like the Kohs Block, Weigl form-color, and Shipley-Hartford were extremely successful in differentiation, and, in combination with the somewhat less accurate EEG, they could discriminate almost perfectly between groups. The tests were used clinically and the specific contribution of each is not presented, unfortunately. Similarly, Hoedemaker and Murray (139) noted almost perfect discrimination of 16 brain-injured from 16 schizophrenics and 16 neurotics by the use of a battery including the Wechsler-Bellevue, Rorschach, Szondi, Ellis Visual Designs, and B.R.L. Sorting tests. Inspection of their breakdown of the various functions tapped reveals that thought process disturbance (tapped chiefly by the sorting tests) and memory were most consistently found and most accurate in differentiating. Electroencephalograms were less accurate, but the test battery combined with EEG and routine neurological examination led to perfect discrimination.

Although originally devised as a nonverbal intelligence test, the Kohs Block test has proven a valuable and interesting technique in diagnosis of brain-injured patients. Wechsler (275) reported it to be one of the most useful subtests in his intelligence

scale for this purpose, while almost all the investigators using the Wechsler-Bellevue scale with brain-injured patients indicate severe impairment on the Block Design subtest (**4, 7, 9, 92, 107, 183**). To the extent that a combination of analytic and synthetic capacities is demanded of the subject by this test, it may measure the primarily abstract components of general intelligence. Thus, a study by Lidz, Gay, and Tietze (**183**) has shown a significant difference in mental age scores obtained from Vocabulary tests and the Stanford-Binet and the mental age scores obtained from the Kohs Blocks. A comparable group of 15 schizophrenics showed no significant differences between the three types of tests.

Significant attempts to devise more elaborate and subtle tests of conceptual or abstraction ability have been reported recently (**103, 113, 120, 121, 141, 241, 264**). A basic need in this area is a far more extensive study of the role of abstract processes in normal human behavior, and studies like those of Heidbreder (**137**) or Hanfmann (**122**) represent only tentative beginnings.

A second major problem in the area of abstract processes following brain lesion has been the question of localization. In the studies summarized by Klebanoff (**164**) the predominant trend suggested that lesions of the frontal areas in particular led to disturbance in abstract behavior. Studies by Rylander (**236**) and Halstead (**120**) particularly have supported this view. The latter investigator presented a small number of cases demonstrating greater loss of abstract ability on the category test in patients with frontal ablations than in patients with lesions in other lobes. Using a study of 147 cases of traumatic head injury, Halstead reported that the tendency these patients showed for impairment, as compared with controls, suggests again the greater effect of frontal injury, since mechanical and autopsy studies (**65, 140**, indicate that frontal injury is most likely to follow any type of head trauma. Lacking neurotic controls for his head injury group, Halstead's results are of limited generality, in view of the findings of Ross and Ross (**231**) which suggest difficulties in differentiating between these groups. The fact is that Halstead's patients with nonfrontal lesions also obtained scores considerably below those of the controls. This suggests impairment which, in view of the small number of subjects, may not be, practically speaking, less significant than findings for the frontal lobe patients. Finally, the age differences of the groups indicate that the frontal lobe patients were, on the average, seven years older than nonfrontal patients, who, in turn, tended to be somewhat older than the controls, so that age differences rather than location of lesion might account for the results. It would appear, therefore, that Halstead's views are in need of further validation.

A series of well-executed studies (**30, 264, 268**) has attempted to test the hypothesis that abstract thinking and fairly complex visual functions are most readily impaired by frontal lesions. These authors studied three well-matched groups of veterans, patients with anterior lobe traumatic lesions, patients with parieto-occipital traumatic lesions, and patients with peripheral nerve injuries who served as nonbrain-injured controls. Lesions were localized by wound of entrance. From a varied series of complex visual tasks, including the Gottschaldt "hidden figures," the

Wisconsin Sorting test (103), a visual-choice reaction test, and an adaptation for humans of Maier's reasoning situation, these authors concluded that test performance of groups of brain-injured was significantly below that of control subjects, despite fairly equal motivation for solution. On some tasks, the occipital lobe patients were inferior to the frontal patients. The authors feel that these results suggest that complex visual performance involving aspects of abstraction is at least as difficult for patients with posterior lesions as for patients with frontal lobe lesions.

While these results cannot be compared directly with those of Halstead (120) because of the difference in the nature of the pathology (trauma versus tumor), they do point up the serious consequences of any type of brain injury, particularly when it is recalled that the patients, unlike most tumor patients, were tested four to seven years after injury. On the other hand, since both the studies of Halstead (120) and Teuber and Bender (268) lack any anatomical or pathological data definitely localizing the injuries, the results have only limited value for theories of brain function. In view of the findings of Holbourn (140) and Courville (65), Halstead might well argue that the trauma of the impact of high velocity missiles in any area of the brain would most likely involve the frontal lobes in any case, thus questioning the conclusions of Teuber and Bender.

An interesting contribution to the problem of localization comes from results presented by McFie and Piercy (192), who studied 74 patients, largely tumor cases, with unilateral lesions. Employing the Weigl form-color sorting test, they found that patients with left-sided lesions proved significantly inferior to patients with lesions located in the so-called nondominant right side, *irrespective of location in the frontal, parietal, or occipital lobes.* This finding prevailed even when the aphasic patients were excluded. These results somewhat contradict those of Anderson (16) whose studies with the Wechsler-Bellevue subtests showed opposite results for patients with dominant and nondominant hemisphere lesions. Here the former patients proved inferior in verbal tests and superior in performance tests. Difficulty in comparing these results, a persistent problem because of brief reporting, makes conclusions necessarily tentative.

Personality Following Brain Lesions and Ablations

Research on personality functioning following brain lesions and ablations has been based chiefly upon the use of the Rorschach test. A major difficulty arises in conclusions drawn from the Rorschach since it is apparent that the critical factors leading to differentiation of organic patients from others derive not from general personality characteristics as much as from disturbances in thinking and perception of the types described above.

The practical usefulness of the Rorschach in studies of brain injury is emphasized in a number of studies. Aita, Reitan, and Ruth (5) compared 60 patients with posttraumatic brain injury with 100 controls representing a heterogeneous group of hospital patients. While quantitative analysis of Rorschach records yielded no consistent picture, use of certain of Piotrowski's qualitative signs (218), e.g., impotence, perplexity, repetition, and color-naming, in addition to

other signs suggesting extreme concreteness, catastrophic reactions, inflexibility, and vagueness, proved helpful. Many neurotic-like signs, depression, anxiety, and hypochondriasis were also found among the organics. Insufficient data are presented for more definitive evaluation of the discriminability of these signs, but it is clear that disturbance in the capacity for abstract thinking manifested by rigidity and extreme concreteness emerged primarily.

Koff (168) similarly reports the usefulness of the Piotrowski (218) signs in differentiating postconcussion neurotics from patients with spinal tap evidence of brain involvement. Again, differentiation is based chiefly upon those aspects of Rorschach performance which may well reflect thought and perceptual disturbances as the primary impairment. Further evidence which supports the value of a related technique calling for subjects to draw impressions from the Rorschach cards comes in a series of papers (51, 104, 179). Using both the Rorschach and the Graphic Rorschach, Grassi (104) reports ten signs generally similar to those of Piotrowski which proved most discriminating between organics and other syndromes. A distinction between blot and concept dominance in performance is made by these authors who point out the extent to which brain-injured patients are blot-dominated.

The general trend of results suggests again primarily primitive thinking and concreteness rather than a basic personality disturbance. Similar indication of a perceptual and thinking impairment is suggested by the work of Hughes (143), who derived a new list of Rorschach signs to differentiate organics from other groups of patients and normals by factor analysis. Another study by Ross and Ross (231) evaluated a number of different clusters of signs and, after extensive manipulation, developed a fairly complex scoring system of "instability" and "disability" ratings that differentiated normals from neurotics and brain-damaged patients.

The "sign" approach, as seen in these studies, seems to work fairly well with brain-injured patients, but extensive cross-validation seems essential. There has been little effort to establish the reliability of various "signs." In general, qualitative analyses of the Rorschach protocols seem to be most successful in selecting brain-injured patients. The sign approach, excessively empirical, gives little real feeling for the basic personality factors that emerge as a consequence of brain injury. It should be noted that signs described, such as impotence, perplexity, inflexibility, etc., are not intrinsic, specifically called forth by the Rorschach blots. They merely represent behavioral evidences of slowed reaction times, abnormal concreteness, perceptual disturbances, and awareness of impaired functioning. It seems that the Rorschach has thus far offered little that is new toward our understanding of the personality of the brain-damaged. However, the value of intensive Rorschach studies of individuals with brain damage is brought out by several case studies (2, 71, 207, 306).

Somewhat unique in the field of personality studies of the brain-injured is a research by Anderson and Honvik (17) comparing Minnesota Multiphasic Personality Inventory profiles of patients with frontal lobe lesions and with parietal involvement. The frontal lobe patients approximated the clinical picture of the

"hysteriod reaction type" while the parietal patients more closely resembled that of the "anxiety neurosis." Since this study lacked any normal or neurotic control groups and since no estimate of the degree of overlap is presented, conclusions remain tentative.

Summarizing the results of the meager personality studies of brain-injured patients, it would appear that in general these patients manifest abnormal concreteness, diffuse and overly generalized modes of organizing their experiences, uncontrollable emotional outbursts, diffuse anxiety ("catastrophic reactions"), and a profound sense of personal inadequacy (1, 5, 104, 148, 155, 179, 306). Lacking in the studies of the personality of the brain-injured has been any attempt to consider the personality dynamics and the adequacy of interpersonal relationships following brain damage of various types. Comparatively little effort at systematically observing social interaction patterns of the brain-injured has been reported, although some authors (6, 94, 203, 278) have referred to problems of this sort.

Psychological Concomitants of Epilepsy

Psychological studies of patients with convulsive seizures diagnosed as epileptics have sought generally to answer two questions: Is there evidence of psychological deficit or deterioration in patients with epileptic seizures? Are there distinctive personality types or patterns of traits characteristic of epileptic patients? Since epilepsy has generally been considered a physiochemical disease (177), treatment has been largely by chemical means. For many years, it was felt that the chemical disturbance in the brain led necessarily to

mental deterioration. By 1943, sufficient data had appeared in a variety of studies to lead Lenox (177) to conclude in his review that mental deterioration was by no means a concomitant of idiopathic epilepsy. A number of studies employing psychometric intelligence tests, usually the Stanford-Binet, failed to find any signs of deterioration (18, 74, 245, 252).

Concern in the past decade has centered not only upon evidence of deterioration but upon exploring characteristic performance patterns of idiopathic epileptics with standard intelligence scales. Collins and Lenox (61) studied a large, heterogeneous group of epileptic outpatients with the Wechsler-Bellevue and found that the group was of better than average intelligence, probably the result of a socioeconomic selective factor in the group studied. A somewhat better controlled study by Sands and Price (239) compared Wechsler-Bellevue patterns of idiopathic epileptics with and without "personality problems." The groups showed average intelligence with only slight differences in their patterns.

Somewhat more definitive findings based on better controlled studies have been reported by Goldman (91) and Winfield (298). The former compared Wechsler-Bellevue patterns of equated groups of idiopathic epileptics, patients with hysterical seizures, and patients with brain lesions but without seizures. In general, the epileptic patients and neurotics showed a similar pattern and both groups differed strikingly from the patients with organic damage, whose performance was generally impaired in the areas of concentration and abstract thinking. No impairment in intellectual functioning was observed in the idiopathic epileptics.

In a similar study Winfield (298) compared intellectual performances of well-equated groups of idiopathic ("cryptogenic") epileptics, symptomatic epileptics with known lesions, posttraumatic brain-injured patients without seizures, and normal controls. Where possible, school IQ or military test scores were used to insure matching of premorbid intelligence levels. Brief tests of verbal meaning, spatial relations, reasoning, associate learning, and abstraction were employed. Results indicated no significant differences between the controls and the idiopathic epileptics, while the two groups with known brain damage showed significantly lower scores in all areas. As in the study of Goldman (91), the seizure symptom seemed less significant than the factor of structural brain damage in relation to intellectual impairment.

In general, results of recent studies of the intellectual performance of idiopathic epileptics do not provide any evidence of intellectual deterioration or of any clear-cut characteristic pattern of performance unique to these patients. The importance of distinguishing between the clearly idiopathic or cryptogenic epilepsies and those who show seizures following traumatic head injury or tumor has been indicated, since in the latter group intellectual impairment does seem to occur.

A more complex problem in the psychological study of epilepsy has been the attempt to delineate a personality pattern characteristic of epileptic patients. Psychological research in this area has largely been restricted to Rorschach studies, following the original descriptions by Rorschach of a number of signs observed in epileptics. Rorschach's signs suggest a general lowered mental functioning level, poor emotional control, and difficulty in abstraction. These findings, partially confirmed by Guirdham (109), may, however, be a function of the use of heterogeneous, institutionalized epileptics, many of whom may have been symptomatic cases. Rorschach studies (19, 125, 169) revealed no peculiarly epileptic patterning, although tendencies toward poor emotional control were frequently observed. Lisansky (185), in a comparison of idiopathic epileptics and diabetics, while yielding no clear-cut pattern, pointed to slower responsivity and a greater evidence of "neurotic signs" among the epileptic population. The findings of these studies agree chiefly in the long response time observed, poor form quality, emotional constriction, and poor emotional control. In view of the heterogeneous nature of the epileptic samples, however, the findings, which strongly resemble those for brain-injured patients on the Rorschach, are difficult to interpret.

The study by Goldman (91) is one of the few employing well-matched groups. Goldman found no clearly defined personality pattern for the idiopathic epileptics. They did, however, show quantitative and qualitative manifestations in the Rorschach of greater drive for achievement, poorly controlled emotionality, immaturity, lowered "inner control and poise," inward turning of affect, and sexual disturbance. In general, the epileptics resembled the hysterical seizure patients more closely than the brain tumor patients in these personality characteristics. These results, while somewhat limited because of the absence of normal controls, point to a strong emotional involvement in epileptics.

PSYCHOSURGERY

Attempts to determine the functions of the frontal lobes in man, based upon studies of individuals with brain pathology, have been hampered by the lack of evaluations of premorbid behavior, the uncontrolled destruction of brain tissue, and the frequent difficulty in localization of areas affected. The various psychosurgical procedures developed may avoid these problems to some extent, but not without many seriously limiting complications. While individuals with presumably anatomically intact brains are subjected to relatively deliberate brain damage, these persons necessarily are in the midst of great emotional upheaval whether they suffer from excruciating pain, severe neuroses, or psychoses. In many cases, preoperative measures of specific functions are not obtainable. Limiting the population to accessible patients introduces a sampling bias of unknown direction and degree. When examination is possible, the reliability of the obtained data, particularly with psychotic patients, is open to question. The assessment of postoperative impairment of changes in functioning is additionally hampered by the fact that many of these persons were impaired intellectually as a result of emotional factors prior to the introduction of a brain lesion.

Essentially, frontal lobe tissue can be rendered nonfunctioning by three methods: (a) direct attack on the thalamus, a technique called thalamotomy or stereoencephalotomy; (b) cutting into the white matter and destroying varying numbers of fibers connecting the cortex and the thalamus—lobotomy, leucotomy, cortical undercutting, etc.; (c) excision or destruction of the cortex—prefrontal lobectomy (215), topectomy (199), gyrectomy (212), and thermocoagulation and venous ligation (200). These diverse techniques have been developed in attempts to improve upon surgical procedures and to avoid some of the undesirable personality changes accompanying the techniques as employed initially. Surgical precision, however, appears to lag far behind the specificity of the hypotheses underlying these variations. In postmortem studies Meyer and McLardy (201) reported great variability in the anatomical lesions produced in lobotomized patients regardless of the intentions of the neurosurgeon. In direct attacks upon the cortex, accuracy generally is not increased because of individual differences in human brains and the lack of clear differentiation among areas. Even in thalamotomy, which in theory appears to be a most precise method, additional damage may result from the gases generated during electrolysis and interference with blood vessels (200). The large variety of psychosurgical techniques plus the possibility of considerable variation in the extent of the lesions produced by means of any one operative procedure contribute to the difficulty of evaluating experimental findings in terms of altered functions.

Many uncontrolled factors are also at play after surgery which obscure and confound changes in functioning. Social improvement or at least increased cooperativeness is a frequently reported concomitant along with altered attitudes of clinical personnel and relatives. Attempts to organize the many and apparently contradictory findings of various investigations are difficult because of inadequate experimental designs. Samples are generally small; adequate control subjects are rarely

available; the time of testing, both pre- and postoperatively, varies considerably within as well as among studies; practice effects are generally ignored; and statistical evaluation of data is frequently omitted or inadequate. In an attempt to bring some structure into this confusion, the present review will deal as much as possible with data from specific techniques rather than with the interpretations made by the various authors.

Standardized Intellectual Tests

Table 1 presents a summary of the data in a number of studies in which standardized intellectual tests (Stanford-Binet and Wechsler-Bellevue) were administered pre- and postoperatively with quantitative data reported. Examination of the postoperative changes reveals that most of these studies show a decrement in IQ for operated patients which in several instances is quite large or statistically significant (172, 197, 214, 238, 304). Nonsignificant changes are reported in the others (191, 199, 200, 220, 262, 272).

Before evaluating these findings, practice effects must be assessed. In the first Columbia-Greystone report (199), Form I of the Wechsler-Bellevue was administered once preoperatively and twice postoperatively. In the second investigation (200), Form I was given twice preoperatively and twice postoperatively and Form II at the fifth session. While both operated and control groups manifested fairly consistent increases in IQ, the operated groups gained less (statistically not significant) than the control groups. The apparent increment in intellectual ability can thus be attributed to other factors, most likely practice. This effect also is evident in Form

II of the Wechsler-Bellevue test, although to a lesser extent than in Form I. The question now arises as to the possibility that the decrements reported in the other studies might be of greater magnitude had control groups been available for comparison. Information concerning possible practice effects are particularly necessary to evaluate the McCullough (191) and Rylander (238) studies, which involved more than one retest. In retesting the same patients nine months postoperatively, Petrie (213) reports that the significant loss in IQ persisted, suggesting that with sufficient time interval between retests, practice effects may be minimized.

Notwithstanding the possible practice effects, all but two of the studies in which data are presented, show some decrement in postoperative performance in comparison to preoperative IQ or control group scores. Some investigators who employed other intelligence test batteries also report postoperative losses in IQ (13, 24, 172). These studies, while not entirely conclusive, strongly suggest that intellectual impairment measurable by means of standardized test techniques does occur following therapeutic destruction of frontal lobe tissue.

With regard to preoperative level and severity of illness, several factors stand out which should be considered in further studies. The higher the preoperative intellectual level, the more likely a large or significant decrease postoperatively (197, 214, 238, 304). In testing a hypothesis dealing with increased variability in the behavior of posttopectomy patients, Wittenborn and Mettler (299) found that ten psychotic patients with relatively high preoperative Wechsler-Bellevue scores tended to

TABLE 1

SUMMARY OF STUDIES PRESENTING PRE- AND POSTOPERATIVE IQ's
FOR PSYCHOSURGERY PATIENTS

Author	N	Diagnostic Category	Surgical Technique	Test	Mean Preoperative IQ	Time of Post-test	Mean IQ Change	Remarks
Porteus & Kepner (220)	18	Psychotic	lobotomy	SB	83.9	varied	−3.2	modified Stanford Binet
Rylander (238)	5	Neurotic	lobotomy	SB	116.2	varied	−11.6*	calculation by the authors
Strom-Olsen et al. (262)	11	Psychotic	lobotomy	SB	94.3	6 wks.	−2.2	calculation by Crow (67)
Yacorzynski et al. (304)	1	Psychotic	lobotomy	SB WB	118 104.5	3 mos. 3 mos.	−21 −17.5	pt. received two preoperative WB's
Koskoff et al. (172)	5	Normals with intractable pain	lobotomy	WB	87.2	3 mos.	−20.4*	
Malmo (197)	6	Neurotic	5 lobotomy 1 gyrectomy	WB	107.8	1–3 mos.	−8.3*	WB Form II on retest
McCullough (191)	10	Psychotic	lobotomy	WB	83.4	2 mos.	2.9	mean IQ change calculated from 2nd postoperative test
Petrie (214)	20	Neurotic	lobotomy	WB	105.8	2–3 mos.	−5.0*	
Vidor (272)	21	Neurotic & Psychotic	lobotomy	WB or SB	111.5	varied	0.6	16 pts. received SB 5 pts. received WB
King (199)		Psychotic	topectomy	WB	101.7	3 wks.	3.9	surgical group gained 3.7 points less than control group
Sheer et al. (200)	20	Psychotic	misc.	WB	79.3	6 mos.	3.4	WB Form II on retest. Surgical group gained 4.5 points less than control group

* Significant at .05 level of confidence.

decrease more than four controls with similar initial scores. Fernandes (76), although giving no data, mentions that patients obtaining higher preoperative scores tended to drop while those with low scores tended to increase. In Table 1, the patients who obtain the higher IQ's are, for the most part, diagnosed as neurotic rather than psychotic. It seems likely that more accurate and valid estimates of intellectual ability can be obtained with neurotics, so that the impairment that occurs following psychosurgery is more readily observed in these patients. The investigation of patients lobotomized for relief from pain, moreover, reveals a large postoperative decrement (172). To what extent the severe pain interfered with optimal functioning is not known, but the postoperative relief apparently was not sufficient to overcome the loss due to brain lesion. A complicating factor in studying such populations is the fact that many of these patients are approaching death rapidly. A nonoperated control group or at least repeated examination seems essential.

With regard to the Verbal and Performance scales of the Wechsler-Bellevue, higher Verbal than Performance losses have been reported (153, 172, 214, 304). The differential effects on Verbal and Performance tasks, however, cannot be evaluated without a control group for comparison. Abilities measured by performance tests may be as impaired as verbal functions following psychosurgery, but the impairment may be obscured by improvement due to differential practice effects. Such improvement also makes it difficult to evaluate the permanence of any changes, although Petrie's study (213) does suggest the possibility of permanent impairment. Social improvement does not seem to bear a relationship to intellectual change as measured by standardized intellectual tests (199, 200).

Miscellaneous Cognitive Functions

In dealing with the more specific aspects of cognitive ability, such as planning ability, abstract ability, memory, learning, attention, etc., the reported findings become considerably more difficult to evaluate because of the variety of tests employed which presumably measure these functions. In examining these functions by means of the specific tests employed rather than the purported specific functions, the confusion is alleviated somewhat, and it is possible to suggest trends.

Porteus Mazes

Despite the susceptibility to practice effects of a performance test such as the Porteus Mazes, impairment below the obtained preoperative level is revealed in many studies on the *first* postoperative examination (197, 199, 200, 214, 220, 221). The time of the first postoperative test varies considerably among and occasionally within studies, but the immediate effects of surgery probably are not involved. Crown (67), reanalyzing Porteus and Kepner's and Porteus and Peter's (220, 221) data, reports that 23 patients whose first postoperative tests took place after three or more months show a highly significant decrement. Preoperative tests were not administered by Robinson (83), but a significant 3.3-year difference in favor of a socially improved nonoperated psychotic control group over the lobotomized group was found. Two studies (153, 262) show insignificant improvement following psychosurgery. The group Jones (153) employed, however, obtained almost a minimal score preoperatively.

A summary of the data of several studies in which the Porteus Mazes were administered pre- and postoperatively is presented in Table 2. Only the changes in the first postoperative examinations are given. Generally, additional examinations result in gains which eventually reach or exceed the preoperative level, but Sheer *et al.* (200) point out that the operated group gained less from practice than did the nonoperated controls. In the first Columbia-Greystone study (199), one year posttopectomy the surgical group had attained the same mental age level as the nonoperated controls. In Petrie's nine months postoperative study (213), the lobotomized patients did not regain the immediate postoperative loss completely, but the difference was no longer significant. Possibly the practice effect was reduced by the six-month interval between tests.

Porteus and Peters (221) and King (199) believe that a pattern of distinct loss on immediate postoperative

TABLE 2

SUMMARY OF STUDIES GIVING PORTEUS MAZES TO PSYCHOSURGERY
PATIENTS PRE- AND POSTOPERATIVELY

Author	N	Diagnostic Category	Surgical Technique	Preoperative Mean MA	Time of Posttest	Mean MA Change	Remarks
Jones (153)	24	Psychotic	lobotomy	5 0	3 wks.	3.0	
Koskoff et al. (172)	3	Normals with intractable pain	lobotomy	not given	3 mos.	−4.1	
Petrie (214)	20	Neurotic	lobotomy	not given	3 mos.	−1 8*	
Porteus & Peters (221), Porteus & Kepner (220)	72	Psychotic	lobotomy	11.3	varied	−1.7*	combined data ana lyzed by Crown (67)
Strom-Olsen et al. (262)	11	Psychotic	lobotomy	12.0	6 wks.	0.1	calculated by Crown (67)
King (199)	19	Psychotic	topectomy	13.2	3 wks.	−1.2	surgical group mean MA is 2.2 yrs. below control
Sheer et al. (200)	23	Psychotic	misc.	10.8	10 days	−1.5	immediate preoperative MA (2nd pretest) is presented

* Significant at or beyond .05 level of confidence.

testing followed by gains up to or beyond the preoperative level in subsequent examinations is related to social recovery. This observation merits further investigation.

Abstract Thinking

Studies which report impairment on standardized intellectual tests also find impaired abstract ability (197, 214, 238, 304). Rylander (238) gained this impression from interpretation of proverbs and fables and the definitions of certain abstract words. Petrie (214) similarly supports this view through the use of the Stanford-Binet proverbs. In addition, however, several studies which report no over-all intellectual impairment do report some decrement in ability to think abstractly (78, 105, 160, 199, 200).

Studies employing the Capps Homograph test consistently report a decrement in the ability to shift. Malmo (197) reports a slight but definite impairment. The first Columbia-Greystone project (199) found a statistically significant decrease ten days posttopectomy. While most patients regained the original loss, the decrease in the ability to make verbal shifts was not regained after one year in six of the 19 subjects. In the second Columbia-Greystone investigation (200) the definite loss was confirmed at the ten-day postoperative test. The loss was regained at the end of 30 days owing to reacquisition of identical definitions originally given but lost postoperatively. Such studies suggest that psychosurgery results in a definite, but probably transient, deficit in verbal ability to shift.

Various sorting, grouping, and

block design tasks are less consistently reported to reveal loss due to frontal lobe surgery. Employing the Goldstein-Scheerer tests with 42 lobotomized patients, Atwell (23) found only slight impairment. Grassi (105) found that patients who showed no general intellectual impairment, evinced a temporary reduction in achievement on his Block Substitution Test. This decrement was almost completely eliminated at the end of a year. Practice effects were not considered. Freudenberg and Robertson (85) report a significant postoperative loss on the Kohs Blocks. The operatees also gained significantly less than did controls on a sorting test. With the modified Kohs Blocks, Weigl color-form, and the Halstead Object-Sorting tests, Kisker (160) concluded that some impairment does occur postoperatively.

While King (199), employing a wide number of measures of abstraction, failed to show any posttopectomy group changes, certain patients did show abstraction difficulties postoperatively which suggested the possibility that such deficit might be related to specific areas of the cortex excised. In the second project (200), postoperative deficit on a modified Weigl test at the ten-day retest was evident, with the loss regained by most patients at the three-month retest.

Some improvement on the color-form sorting test is reported by Jones (153), but no controls were used to determine the amount of gain attributable to practice. Hunt (81) and Strom-Olsen et al. (262) report no change in pre- and posttesting with the Kohs Blocks while studies employing the Shipley-Hartford consistently report no impairment postoperatively (14, 79, 228, 262).

Memory, Learning, Attention

Although memory and related functions have been found to be extremely sensitive to impairment associated with brain pathology in clinical situations, the reported investigations of individuals undergoing psychosurgery have not been fruitful. The gross attention and concentration difficulties characteristic of emotionally disturbed persons probably interfere markedly with preoperative assessment of these functions. Some studies simply report no gross or permanent changes postoperatively (13, 14, 194, 238). Extremely varied results on rote memory tasks are reported by many investigators (24, 153, 191, 197, 228). Thus, Freudenberg and Robertson (85) report significant memory loss postoperatively in a group of 24 lobotomized patients on paired associates and recall of the Bender-Gestalt figures, but no impairment in memory for objects, perhaps because of the greater concreteness of the latter task. A fairly comprehensive investigation of memory and learning was undertaken by Stauffer (199). Control and topectomy groups learned semimeaningful and meaningful paired-associate lists and a paragraph of verbal directions preoperatively. In general, retention of previously learned material and the ability to learn new material was unaffected by topectomy.

King (199) also included in his study of intellectual functions, a continuous-problem task which utilized an instrument intended for selection of pilots by the Army Air Forces. The test involved a complicated choice reaction which permitted obtaining a measure of "the patient's ability to perform on a task requiring close attention and sustained effort

over a period of time." While the topectomy group suffered no marked impairment on this task, the control group manifested a greater trend in the direction of more problems solved and fewer errors than did the operated group. Robinson (**83, 228**), testing the capacity for prolonged attention and deliberation by means of simple arithmetic problems, rhyming, and three of the Downey Will-Temperament Tests, reports clear-cut deficit in the lobotomized group as compared to nonoperated controls. Malmo (**197**) also found that deliberation was reduced postoperatively as indicated by more rapid performance on Raven's Progressive Matrices Test. Reduction in time and score on the Matrices Test as well as very rapid performance on the MMPI postoperatively was observed by Vidor (**272**). In the Columbia-Greystone project the topectomized group showed a trend toward poorer performance in addition tests, but not significantly different from the controls. A subtraction test showed significant changes in variability attributable to the operation. On a cancellation test, Rylander (**238**) reports a decrease one month postoperatively with subsequent gains to preoperative level. Hunt (**81**), however, reports postoperative increased time with improvement in accuracy on a cancellation test, as well as improvement in immediate memory.

Sensory Functions

The relief of intractable pain by means of psychosurgery is apparently afforded without increasing the threshold for pain. Using ordinary clinical tests in neurological examination of such patients, Watts and Freeman (**274**) failed to disclose any evidence of impaired sensation. Indeed,

Chapman, Rose, and Solomon (**55**) report increased withdrawal reactions to pin prick are not uncommon in lobotomized patients. Following up this clinical impression and employing the Hardy-Wolff-Goodell apparatus for more accurate evaluation, they found that postoperatively their group of 23 psychiatric patients withdrew their heads from the apparatus at less intense levels of stimulation than before lobotomy. Over a two-year period, consistent trends toward a return to preoperative threshold levels were revealed (**56**). Malmo (**197**) also found a much higher rate of withdrawal after lobotomy. King *et al.* (**158**) supported these findings in a study of five patients operated unilaterally for relief of pain. In the second Columbia-Greystone study (**200**), slightly reduced thresholds were observed in only two of the nine operated patients. In this study, the radiant heat was applied to the forearm rather than the forehead.

Extensive psychophysiological studies were employed in both Columbia-Greystone projects. No marked losses or consistent changes were found in auditory acuity, visual acuity, peripheral vision, brightness discrimination, color vision, time judgment, autokinetic effect, critical flicker frequency, recognition of tachistoscopically exposed stimuli, or various motor tasks.

Personality

In his review of the clinical studies of lobotomized patients Crown (**67**) points out that "almost invariably . . . *personality changes* have followed the operation." These changes are in the direction of increased cheerfulness, complacency, apathy, restlessness, shallowness of affect, indifference to criticism and feelings of others, and decreased self-conscious-

ness, reserve, and tact. Kolb (171) also points out that there is little difference of opinion among clinicians as to the nature of the personality changes following lobotomy.

Rating scales have been employed in several studies (153, 226, 244), which generally found improved ward behavior. Cooperativeness, sociability, and tidiness are increased; anxiety, depression, and bizarre behavior are decreased. Affect and feeling are adversely affected in the direction of greater apathy (226). Lack of initiative is also noted (244). In a series of ratings, Jones (153) indicates that for the group, over-all behavior did not improve after the eighth week post-operatively. A time sample of behavior was obtained for surgical patients alone and in pairs by Kinder and Willenson (200) with no changes in patterns of behavior apparent following surgery. Patients under 40 showed some increase in activity while those over 40 manifested a decrease. Bockoven and Hyde (106) observed 16 patients pre- and post-lobotomy in groups, recording sociograms of the patients' interactions. Improvement in psychiatric cases was generally accompanied by increased socialization and development of a friendly democratic attitude toward other patients. It is possible that increased familiarity with observers and examiners and the additional attention may lead to improvement and reduction in anxiety.

The necessity for control groups is suggested by the findings of Wittenborn and Mettler (299). Employing symptom rating scales, no significant differences between the topectomized and control groups were found; both groups manifested a reduction in symptomatology. Indeed, a return of certain pathological symptoms was noted in some surgical patients who had been free from such pathology prior to topectomy.

Noticeable lessening of psychotic symptomatology is reported for lobotomized patients in studies employing personality inventories (14, 81, 199, 272, 304). Standardized interviews utilized to obtain attitudes dealing with feelings of guilt, religion, sex, and prejudice failed to reveal any over-all attitudinal changes (200). Employing a sensibility questionnaire to measure degree of concern with one's past and future and with the opinions of others and a self-regarding span (a measure of the time spent talking about oneself), Robinson (83) compared a lobotomized group with a nonoperated control group and found that the operated patients showed less self-preoccupation, less concern with the past, future, and opinions of others. Thus, for the most part the feelings and attitudes expressed by patients undergoing psychosurgery are in accord with the clinical observations and ratings made of their behavior.

In summarizing the Rorschach data for the first Columbia-Greystone project, Zubin (199) points out that some subjects showed altered personality trends, but that no definite patterns of changes emerged. Analysis of group changes revealed a pronounced decline in reaction time for the operated patients. Suggested trends were posttopectomy decreases in number of responses and factors primarily associated with anxiety, ambitiousness, conflict, introspection, and perceptual accuracy. Atwell (23) also reports increased constriction, perseveration, and stereotypy along with less spontaneity, initiative, and fantasy. The most marked postoperative change in Atwell's study was the patient's carefree and uncon-

cerned approach to the Rorschach. Similar changes in Rorschach factors also emerged in studies employing fewer subjects (86, 304). Jones (153) found an increase in responses, but taking into account the initially constricted record and possible practice effects of repeated testing, these results do not contradict the above-mentioned studies.

Some retest changes apparently may occur regardless of surgery according to Wittenborn and Mettler (299). Employing a novel Rorschach measure (lack-of-perceptual-control score), revealed by responses in which form is absent or secondary to color or shading, they found a significant difference between topectomized patients and controls. Topectomy patients increased in "lack of perceptual control" postoperatively while control patients showed a decrease in this measure.

In a comparison of pre- and post-lobotomy Rorschach records of 40 patients, Hunt (81) reports greater constriction two weeks postoperatively. Somewhat contradictory to the other studies are findings of increased populars and testing time and decreased perseveration and self-references. The patients also manifested less reluctance in their approach to the task, less self-criticism and concern over performance. Amaral (13) reports an almost identical picture postoperatively for 18 patients.

In a study on creative ability, Hutton and Bassett (147) indicate that the Rorschach, the Harrower-Erickson Multiple-Choice Rorschach, a story-telling, and a drawing test reveal a lessening of creative ability in leucotomized patients. Few patients, however, manifested a creative urge prior to the operation. In a later study, Ashby and Bassett (21) employed a drawing test and found no difference between operatees and psychotic controls. Both patient groups did worse than normal control subjects.

Ashby and Bassett (22) studied the psychogalvanic response of 21 lobotomized patients and 21 controls to real and symbolic (unfulfilled) threat. Six patients were also studied pre- and postoperatively. No uniform postoperative trends emerged; the symbolic threats retained the power of eliciting the psychogalvanic response. Thus, the authors conclude that emotional drive is not diminished by lobotomy.

Petrie (214) reports a postoperative decrease in neuroticism as manifested by reduced body sway suggestibility and a smoother work curve in a neurotic group. Diminished introversion was also evident as revealed by loss in persistence, a tendency to go for speed rather than accuracy, reduced self-blame, greater reality adjustment on level-of-aspiration tests, and a tendency to live in the present.

In general, personality studies do suggest a decrease in depression and anxiety following psychosurgery. These changes seem to occur at the expense of greater personality constriction, decreased critical standards and regard for others. Evaluation of the presence of "organic" indicators is obscured by their presence in preoperative Rorschach records (161, 271) or by the overlap with postoperative psychotic residuals.

CONCLUSIONS

The last decade has witnessed a significant change in the orientation of research in the field of organic brain dysfunction. Klebanoff (164), in his earlier review, had concluded that psychological studies in this field had been concerned primarily

with relating mental functions to localized areas of the brain, the determination of the presence of brain pathology on the basis of psychological test performance, and the detention of an "organic psychological syndrome." Methodologically, essential reliance had been placed upon the use of omnibus test techniques such as the Stanford-Binet and Rorschach. However, the beginning of a trend toward the use of more specialized test techniques had been observed.

The period covered by the present review reveals decreased emphasis upon localization and diagnosis and a concomitant increase of interest in the developmental aspects of brain injury, emphasis upon patterns of functioning through the use of specialized tests including laboratory methods, and an overly hasty concentration of effort upon the psychological consequences of psychosurgery.

During the past ten years, increased theoretical sophistication and research findings have served to dispel the earlier optimism concerning the ability of psychological test techniques to contribute toward localization of brain pathology. The correlation of psychological test performance with specific areas of brain damage has been found to be limited by vast differences in brain pathology caused by different types of injury or disease as well as by serious limitations in the techniques of anatomical localization. In general, psychological instruments have proven incapable of differentiating patients with presumptive injury to specific cortical areas. Indeed, the limited number of studies which do report such differentiation merit repetition or cross-validation, an apparent constant necessity in this field of research.

The continued development of specialized test methods reflects the inability of conventional psychometric techniques to reveal clear pictures of organic impairment. Thus, the considerable research employing the Wechsler-Bellevue scale reveals that although the test as a whole proves relatively insensitive for diagnostic purposes in patients with brain injury, certain of the subtests appear to be quite discriminating in numerous studies. In addition, the trend toward the use of laboratory perceptual techniques, such as the critical flicker fusion, the phi phenomenon, and the tachistoscopic presentation of stimuli, have opened additional horizons for study and merit concerted research exploration. These methods do reveal promise, but one is not yet able to evaluate their ultimate significance as differentiating techniques.

A further significant development in recent years has been the rather extensive research upon children with organic brain disease. Analysis of the deficit findings in children with brain damage reveals that the impairment tends generally to parallel that observed in adults with brain damage. There is, however, the definite suggestion that children with brain pathology manifest marked unevenness and inconsistency in the development of their intellectual capacities, and this may indicate more generalized cerebral localization of function in children. The research directed toward differentiating endogenous and exogenous feebleminded children has suggested promising new psychological approaches and techniques in addition to making interesting theoretical contributions.

Finally, it is felt that subsequent research in this field should recognize the importance of the interaction of such related variables as environment

and premorbid personality of patients with organic brain damage. It is apparent that intellectual deficit must be evaluated in relation to the richness and complexity of past and present environmental situations. Indeed, the tempo and degree of so-called deterioration in certain senile and organic conditions may be determined significantly by the nature of social and other environmental factors. In addition, numerous questions arise that emphasize the need for additional research dealing with the matter of personality changes believed to be associated with organic brain disease. For example, do existing intellectual limitations. present the appearance of fundamental personality alteration or are there basic qualitative changes in the premorbid drives, motives, and basic personality dynamics of patients following organic brain insult?

It is unfortunate that most studies of personality changes in organic brain disease have utilized the Rorschach test and reliance upon test signs which have not been adequately cross-validated. A more fruitful evaluation might involve an approach aimed at an analysis of the particular social, familial, and personal demands made upon the patient with organic brain damage. Following such an analysis of expected and desired behavior, specialized psychological techniques might be employed to evaluate the capacities required to fulfill these environmental demands. Such an approach would appear to offer extremely vital information concerning the consequences of brain injury or ablation. At the present time, there is a striking absence of psychological research designed to understand the altered dynamic field of the adult with brain injury. Similarly, there is a need for scientific investiga-

tion of the relationship between such variables as premorbid personality and socioeconomic and cultural milieu in relation to mode of adjustment to brain disease.

When one considers the literature dealing with psychosurgery, the results prove generally to be disconcerting. Although inconsistent and paradoxical results are observed when identical test techniques are employed, there are some trends in the results reflected in Tables 1 and 2. The varied conclusions in this area of research appear to be a consequence of numerous kinds of errors. First, psychosurgery is an extremely broad term and subsumes a large number of different surgical procedures involving varying degrees of destruction of brain tissue. Second, when psychometric and psychological test techniques are utilized with severely psychotic patients, the reliability or representativeness of the results may be questioned. Third, the results in the majority of the studies are not comparable since different or heterogeneous diagnostic groups of patients were employed. Finally, a large number of studies dealing with the effects of psychosurgery have been marked by faulty experimental design, particularly by absence of the use of adequate control groups.

Despite the difficulties cited above, it is possible to extract some generalizations regarding the impact of therapeutic psychosurgery. For example, in those patients whose preoperative test performance was not markedly disturbed, there is some suggestion of impairment in general intelligence, abstract thinking ability, memory functioning, learning ability, and sustained attention. Qualitative evaluation of personality changes appears to indicate a more apathetic, less complex, and constricted individual

who shows less introspective concern with himself and less depression of mood following lobotomy. It is clear, however, that experimental knowledge of the sequelae of psychosurgery remains restricted and conflicting. Continued research is necessary in this field with subsequent emphasis upon experimental design and the use of control groups.

The present review has attempted to organize and integrate the extensive literature dealing with organic brain damage over the past decade. The voluminous body of literature covered reflects adequately the degree of scientific preoccupation with this area of research. It appears that the change in emphasis from mere diagnosis and localization in the direction of the study of related variables represents a scientifically healthy reorientation. The introduction of laboratory test methods represents significant continued exploration of the utility of specialized techniques. It is felt, finally, that future research upon the patient with organic brain disease should regard him as a complex individual whose social, economic, and intellectual environmental demands must be considered in order to attain total understanding of the specific consequences of brain pathology.

REFERENCES

1. Abbot, W. D., Due, F. O., & Nosik, W. A. Subdural hematoma and effusion as a result of blast injuries. *J. Amer. med. Ass.*, 1943, **121**, 739–741.

2. Ackerly, S. S., & Benton, A. L. Report of case of bilateral frontal lobe defect. *Res. Publ. Ass. nerv. ment. Dis.*, 1950, **27**, 479–504.

3. Acklesberg, S. B. Vocabulary and mental deterioration in senile dementia. *J. abnorm. soc. Psychol.*, 1944, **39**, 393–406.

4. Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain injury. *J. gen. Psychol.*, 1947, **37**, 25–44.

5. Aita, J. A., Reitan, R. M., & Ruth, Jane M. Rorschach's test as a diagnostic aid in brain injury. *Amer. J. Psychiat.*, 1947, **103**, 770–779.

6. Alexander, L. The element of psychotherapy in the treatment of organic neurologic disorders. *J. nerv. ment. Dis.*, 1951, **114**, 283–306.

7. Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225–230.

8. Allen, R. M. A note on the use of the Bellevue-Wechsler Scale Mental Deterioration Index with brain-injured patients. *J. clin. Psychol.*, 1948, **4**, 88–89.

9. Allen, R. M. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, **4**, 281–284.

10. Allen, R. M. A comparison of the test performance of the brain-injured and the brain-diseased. *Amer. J. Psychiat.*, 1949, **106**, 195–198.

11. Allen, R. M. An analysis of the comparative evaluation of Allen's brain-injured patients and of normal subjects. *J. clin. Psychol.*, 1949, **5**, 422–423.

12. Allen, R. M., & Krato, J. C. The test performance of the encephalopathic. *J. ment. Sci.*, 1949, **95**, 369–372.

13. Amaral, M. A. Comparative results with Moniz's prefrontal leucotomy and Freeman's lobotomy. *First int. Cong. Psychosurg., Lisbon*, 1949, 173–184.

14. Anderson, A. L. Personality changes following prefrontal lobotomy in a case of severe psychoneurosis. *J. consult. Psychol.*, 1949, **13**, 105–107.

15. Anderson, A. L. The effect of laterality localization of brain damage on Wechsler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1950, **6**, 191–194.

16. Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1951, **7**, 149–153.

17. Anderson, A. L., & Hanvik, L. J. The psychometric localization of brain lesions; the differential effect of frontal and parietal lesions on MMPI profiles. *J. clin. Psychol.*, 1950, **6**, 177–180.

18. ARIEFF, A. J., & YACORZYNSKI, G. Deterioration of patients with organic epilepsy. *J. nerv. ment. Dis.*, 1942, 96, 49-55.

19. ARLUCK, E. W. A study of some personality characteristics of epileptics. *Arch. Psychol.*, 1941, No. 263.

20. ARMITAGE, S. G. An analysis of certain psychological tests used for the evaluation of brain injury. *Psychol. Monogr.*, 1946, 60, No. 1 (Whole No. 277).

21. ASHBY, W. R., & BASSETT, M. The effect of leucotomy on creative ability. *J. ment. Sci.*, 1949, 95, 418-430.

22. ASHBY, W. R., & BASSETT, M. The effect of prefrontal leucotomy on the psychogalvanic response. *J. ment. Sci.*, 1950, 96, 458-469.

23. ATWELL, C. R. Psychometric changes after lobotomy. *J. nerv. ment. Dis.*, 1950, 111, 165-166.

24. BABCOCK, HARRIET. A case of anxiety neurosis before and after lobotomy. *J. abnorm. soc. Psychol.*, 1947, 42, 466-472.

25. BALDWIN, M. V. A clinico-experimental investigation into the psychological aspects of multiple sclerosis. *J. nerv. ment. Dis.*, 1952, 115, 299-343.

26. BARNES, T. C. Electroencephalographic validation of the Rorschach, Hunt, and Bender-Gestalt tests. *Amer. Psychologist*, 1950, 5, 322. (Abstract)

27. BATTERSBY, W. S. Critical flicker frequency in patients with cerebral lesions. *Amer. Psychologist*, 1950, 5, 271-272. (Abstract)

28. BATTERSBY, W. S. The regional gradient of critical flicker frequency after frontal or occipital lobe injury. *J. exp. Psychol.*, 1951, 42, 59-68.

29. BATTERSBY, W. S., BENDER, M. B., & TEUBER, H. L. Effects of total light flux on critical flicker frequency after frontal lobe lesion. *J. exp. Psychol.*, 1951, 42, 135-142.

30. BATTERSBY, W. S., TEUBER, H. L., & BENDER, M. B. Problem-solving behavior in men with frontal or occipital brain injuries. *Amer. Psychologist*, 1951, 7, 264-265. (Abstract)

31. BENDER, LAURETTA. Psychological principles of the Visual Motor Gestalt Test. *Trans. N. Y. Acad. Sci.*, 1949, 11, 164-170.

32. BENDER, LAURETTA. Psychological problems of children with organic disease. *Amer. J. Orthopsychiat.*, 1949, 19, 404-415.

33. BENDER, LAURETTA, & SILVER, A. Body image problems of the brain-damaged child. *J. soc. Issues*, 1948, 4, 84-89.

34. BENDER, M. B., & FURLOW, L. T. Visual disturbances produced by bilateral lesions of the occipital lobes with central scotomas. *Arch. Neurol. Psychiat.*, 1945, 53, 165-170.

35. BENDER, M. B., SHAPIRO, M. S., & TEUBER, H. L. Allesthesia and disturbance of body schema. *Arch. Neurol. Psychiat.*, 1949, 62, 222-236.

36. BENDER, M. B., & TEUBER, H. L. Phenomena of fluctuation, extinction, and completion in visual perception. *Arch. Neurol. Psychiat.*, 1946, 55, 627-658.

37. BENDER, M. B., & TEUBER, H. L. Ring scotoma and tubular fields: their significance in cases of head injury. *Arch. Neurol. Psychiat.*, 1946, 56, 200-226.

38. BENDER, M. B., & TEUBER, H. L. Spatial organization of visual perception following injury to the brain. *Arch. Neurol. Psychiat.*, 1947, 58, 721-739; 1948, 59, 39-62.

39. BENDER, M. B., & TEUBER, H. L. Disorders in the visual perception of motion. *Trans. Amer. neurol. Ass.*, 1948, 73, 191-193.

40. BENDER, M. B., & TEUBER, H. L. Disorders in visual perception following cerebral lesions. *J. Psychol.*, 1949, 28, 223-233.

41. BENDER, M. B., & TEUBER, H. L. Psychopathology of vision. In E. A. Spiegel (Ed.), *Progress in neurology and psychiatry.* New York: Grune & Stratton, 1949. Pp. 163-192.

42. BENSBERG, G. J. A test for differentiating endogenous and exogenous mental defectives. *Amer. J. ment. Def.*, 1950, 54, 502-506.

43. BENTON, A. L., & COLLINS, NANCY T. Visual retention test performance in children, normative and clinical observations. *Arch. Neurol. Psychiat.*, 1949, 62, 610-617.

44. BENTON, A. L., & HOWELL, L. I. The use of psychological tests in the evaluation of intellectual function. *Psychosom. Med.*, 1941, 3, 138-151.

45. BERLINER, F., MAYER-GROSS, W., BEVERIDGE, R. L., & MOORE, J. N. P. Prefrontal leucotomy: report on 100 cases. *Lancet*, 1945, 249, 325-328.

46. BIJOU, S., & WERNER, H. Language analysis in brain injured and nonbrain injured mentally deficient children. *J. genet. Psychol.*, 1945, 66, 239-254.

47 BLAKE, R., & McCARTY, B. A comparative evaluation of the Bellevue-Wechsler Mental Deterioration Index distributions of Allen's brain-injured patients and normal subjects. *J. clin. Psychol.*, 1948, 4, 415–418.

48 BLATT, B., & HECHT, I. The personality structure of the multiple sclerosis patient as evaluated by the Rorschach psychodiagnostic technique. *J. clin. Psychol.*, 1951, 7, 341–344.

49. BOTWINICK, J., & BIRREN, J. E. The measurement of intellectual deterioration in senile psychosis and psychosis with cerebral arteriosclerosis. *Amer. Psychologist*, 1950, 5, 364–365. (Abstract)

50. BRIDGMAN, O. A case study of gross brain damage. *Amer. J. ment. Def.*, 1941, 46, 195–197.

51. BRUSSEL, J. A., GRASSI, J. R., & MELNIKER, A. A. The Rorschach method and postconcussion syndrome. *Psychiat. Quart.*, 1942, 16, 707–743.

52. BUSEMANN, A. Demenz als Dauerfolge von Hirnverletzungen. *Schweis. Z. Psychol. Anwend.*, 1950, 9, 119–128.

53. CANTER, A. H. Direct and indirect measures of psychological deficit in multiple sclerosis. *J. gen. Psychol.*, 1951, 44, 3–50.

54. CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, 40, 153–193.

55. CHAPMAN, W. P., ROSE, A. S., & SOLOMON, H. C. Measurement of heat stimulus producing motor withdrawal reaction in patients following frontal lobotomy. *Res. Publ. Assoc. nerv. ment. Dis.*, 1948, 27, 754–768.

56. CHAPMAN, W. P., ROSE, A. S., & SOLOMON, H. C. A follow-up study of motor withdrawal reaction to heat discomfort in patients before and after frontal lobotomy. *Amer. J. Psychiat.*, 1950, 107, 221–224.

57. CLEVELAND, S. E., & DYSINGER, D. W. Mental deterioration in senile psychosis. *J. abnorm. soc. Psychol.*, 1944, 39, 368–372.

58. COHEEN, J. Disturbances in time discrimination in organic brain disease. *J. nerv. ment. Dis.*, 1950, 112, 121–129.

59. COHEN, J. Wechsler Memory scale performance of psychoneurotic, organic, and schizophrenic groups. *J. consult. Psychol.*, 1950, 14, 371–375.

60. COLE, E. M., BAGGETT, MIRIAM P., & MacMULLEN, MARJORIE R. Mental and performance testing of neurologic patients. *Arch. Neurol. Psychiat.*, 1947, 58, 104–107.

61. COLLINS, A. L., & LENNOX, W. G. The intelligence of 300 private epileptic patients. *Res. Publ. Ass. Res. nerv. ment. Dis.*, 1947, 26, 583–603.

62. COLM, HANNA. The value of projective methods in the psychological examination of children: the Mosaic Test in conjunction with the Rorschach and Binet tests. *Rorschach Res. Exch.*, 1948, 12, 216–237.

63. COLOM, G. A., & LEVINE, M. H. Self-inflicted prefrontal lobotomy: report of a case. *J. nerv. ment. Dis.*, 1951, 113, 430–436.

64. COTTON, C. B. A study of reactions of spastic children to certain test situations. *J. genet. Psychol.*, 1941, 58, 27–35.

65. COURVILLE, C. B. Coup-contrecoup mechanism of cranio-cerebral injuries: some observations. *Arch. Surg.*, 1942, 45, 19–43.

66. CRITCHLEY, M. The body-image in neurology. *Lancet*, 1950, 1, 335–340.

67. CROWN, S. Psychological changes following prefrontal leucotomy: a review. *J. ment. Sci.*, 1951, 97, 49–83.

68. DIERS, W. C., & BROWN, C. C. Psychometric patterns associated with multiple sclerosis. I. Wechsler-Bellevue patterns. *Arch. Neurol. Psychiat.*, 1950, 63, 760–765.

69. DI NOLFO, A. A simple screening device for and in detection of brain damage. *J. Amer. osteop. Ass.*, 1947, 47, 244–247.

70. DOLPHIN, J. E., & CRUICKSHANK, W. N. The figure background relationship in children with cerebral palsy. *J. clin. Psychol.*, 1951, 7, 228–231.

71. ELONEN, ANNA S., & KORNER, ANNE-LIESE, F. Pre- and post-operative psychological observations on a case of frontal lobectomy. *J. abnorm. soc. Psychol.*, 1948, 43, 532–543.

72. EYSENCK, M. D. A study of certain qualitative aspects of problem solving behaviour in senile dementia patients. *J. ment. Sci.*, 1945, 91, 337–345.

73. EYSENCK, M. D. An exploratory study of mental organization in senility. *J. Neurol. Psychiat.*, 1945, 8, 15–21.

74. FALK, R., PENROSE, L. S., & CLAR, S. The search for intellectual deterioration among epileptic patients. *Amer. J. ment. Def.*, 1945, 49, 469–471.

75. FELDMAN, F., & CAMERON, D. E. Speech in senility. *Amer. J. Psychiat.*, 1944, 101, 64–67.

76. FERNANDES, B., *et al.* A clinical and psychological study in leucotomy. *First int. Congr. Psychosurg.*, Lisbon, 1949, 147–165.

77. FISCHER, LISELOTTE K. A new psychological tool in function: preliminary clinical experience with the Bolgar-Fischer World Test. *Amer. J. Orthopsychiat.*, 1950, 20, 281–292

78. FLEMING, G. W. T. H. Some preliminary remarks on prefrontal leucotomy. *J. ment. Sci.*, 1942, 88, 282–284.

79. FLEMING, G. W. T. H. Prefrontal leucotomy. *J. ment. Sci.*, 1944, 90, 486–500.

80. FRANK, J. Clinical survey and results of 200 cases of prefrontal leucotomy. *J. ment. Sci.*, 1946, 92, 497–508.

81. FREEMAN, W., & WATTS, J. W. *Psychosurgery.* Springfield, Ill.: Charles C Thomas, 1942.

82. FREEMAN, W., & WATTS, J. W. Psychosurgery: an evaluation of two hundred cases over seven years. *J. ment. Sci.*, 1944, 90, 532–537.

83. FREEMAN, W., & WATTS, J. W. *Psychosurgery.* (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1950.

84. FRENCH, L. A. Psychometric testing of patients who had brain tumors removed during childhood. *J. Neurosurg.*, 1948, 5, 173–177.

85. FREUDENBERG, R. K., & ROBERTSON, J. P. S. Investigation into intellectual changes following prefrontal leucotomy. *J. ment. Sci.*, 1949, 95, 826–841.

86. FURTADO, D., RODRIGUES, M., MARGUES, U., ALVINS, F., & DE VASCONCELOS, A. Personality changes after lobotomy. *First int. Cong. Psychosurg.*, Lisbon, 1949, 35–49.

87. GELB, A., & GOLDSTEIN, K. *Psychologische Analysen hirnpathologischer Fälle.* Leipzig: Barth, 1920. (Partially translated in Ellis, W., *A source book of Gestalt psychology.* New York: Harcourt, Brace, 1938.)

88. GILLILAND, A. R., WITTMAN, P., & GOLDMAN, M. Patterns and scatter of mental abilities in various psychoses. *J. gen. Psychol.*, 1943, 29, 257–260.

89. GLIK, E. E. A comparison of recall and recognition types of measurement on verbal items, and their implications for deterioration testing. *J. clin. Psychol.*, 1951, 7, 157–162.

90. GOLDENSOHN, L. N., CLARDY, E. R., & LEVINE, K. N. Schizophrenic-like reactions in children. *Psychiat. Quart.*, 1945, 19, 572–604.

91. GOLDMAN, G. D. A comparison of the personality structures of patients with idiopathic epilepsy, hysterical convulsions, and brain tumors. Paper read at East. Psychol. Ass., Atlantic City, 1952.

92. GOLDMAN, R., GREENBLATT, M., & COON, G. P. Use of the Bellevue-Wechsler scale in clinical psychiatry with particular reference to cases with brain damage. *J. nerv. ment. Dis.*, 1946, 104, 144–179.

93. GOLDSTEIN, K. *The organism.* New York: American Book Co., 1939.

94. GOLDSTEIN, K. *Human nature in the light of psychopathology.* Cambridge, Mass.: Harvard Univer. Press, 1940.

95. GOLDSTEIN, K. *After-effects of brain injuries in war.* New York: Grune & Stratton, 1942.

96. GOLDSTEIN, K. Some experimental observations concerning the influence of colors on the function of the organism. *Occup. Ther.*, 1942, 21, 147–151.

97. GOLDSTEIN, K. Brain concussion: evaluation of the after-effects by special tests. *Dis. nerv. Syst.*, 1943, 4, 325–334.

98. GOLDSTEIN, K. Mental changes due to frontal lobe damage. *J. Psychol.*, 1944, 17, 187–208.

99. GOLDSTEIN, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.

100. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).

101. GOMEZ DEL CARRO, J. El test del arbol in clinica psyquiatrica; Koch's Baumtest. (The tree test in the psychiatric clinic: Koch's tree-test). *Acta med. hispanica*, 1950, 8, 53–59.

102. GRAHAM, F. K., & KENDALL, B. S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, 41, 303–314.

103. GRANT, A. D., & BERG, Esta A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting. *J. exp. Psychol.*, 1948, 38, 404–411.

104. GRASSI, J. R. The Graphic Rorschach as a supplement to the Rorschach in the diagnosis of organic intracranial le-

sions. *Psychiat. Quart. Suppl.*, 1947, 21, 312–327.

105. GRASSI, J. R. Impairment of abstract behavior following prefrontal lobotomy. *Psychiat. Quart.*, 1950, 24, 74–88.

106. GREENBLATT, M., ARNOT, R., & SOLOMON, H. C. (Eds.) *Studies in lobotomy.* New York: Grune & Stratton, 1950.

107. GREENBLATT, M., GOLDMAN, R., & COON, G. P. Clinical implications of the Bellevue-Wechsler test (with particular reference to brain damage cases). *J. nerv. ment. Dis.*, 1946, 104, 438–442.

108. GREENBLATT, M., LEVINE, S., & ATWELL, C. R. Comparative value of electroencephalogram and abstraction tests in diagnosis of brain damage. *J. nerv. ment. Dis.*, 1945, 102, 383–391.

109. GUIRDHAM, A. The Rorschach test in epileptics. *J. ment. Sci.*, 1935, 81, 870–893.

110. GUTMAN, BRIGETTE. The application of the Wechsler-Bellevue scale in the diagnosis of organic brain disorders. *J. clin. Psychol.*, 1950, 6, 195–198.

111. HAFFTER, C. Der Labyrinth-Test von Rey bei Oligophrenen, Epileptikern und Organisch Dementen. (Rey's maze test in mental defectives, epileptics, and organic dements). *Mschr. Psychiat. Neurol.*, 1942, 106, 1–10.

112. HALL, K. R. L. The testing of abstraction, with special reference to impairment in schizophrenia. *Brit. J. med. Psychol.*, 1951, 24, 83–150.

113. HALL, K. R. L. Conceptual impairment in depressive and organic patients of the pre-senile age group. *J. ment. Sci.*, 1952, 98, 257–264.

114. HALL, K. R. L., & CROOKES, T. G. Studies in learning impairment. I: Schizophrenic and organic patients. *J. ment. Sci.*, 1951, 97, 729–737.

115. HALSTEAD, H. Mental tests in senile dementia. *J. ment. Sci.*, 1944, 90, 720–726.

116. HALSTEAD, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. *Amer. J. Psychiat.*, 1940, 96, 1263–1294.

117. HALSTEAD, W. C. A power factor (P) in general intelligence: the effect of brain injuries. *J. Psychol.*, 1945, 20, 57–64.

118. HALSTEAD, W. C. Brain injuries and higher levels of consciousness. *Res.*

Publ. Ass. nerv. ment. Dis., 1945, 24, 480–506.

119. HALSTEAD, W. C. A power factor (P) in general intelligence: effects of lesions of the brain. *Arch. Neurol. Psychiat.*, 1946, 56, 234–235.

120. HALSTEAD, W. C. *Brain and intelligence.* Chicago: Univer. of Chicago Press, 1947.

121. HALSTEAD, W. C., & SETTLAGE, P. H. Grouping behavior of normal persons and of persons with lesions of the brain. *Arch. Neurol. Psychiat.*, 1943, 49, 489–506.

122. HANFMANN, EUGENIA. A study of personal patterns in intellectual performance. *Charact. & Pers.*, 1941, 9, 315–325.

123. HANFMANN, EUGENIA, RICKERS-OVSIANKINA, MARIA, & GOLDSTEIN, K. Case Lanuti: extreme concretization of behavior due to damage of the brain cortex. *Psychol. Monogr.*, 1944, 57, No. 4 (Whole No. 264).

124. HANVIK, L. J., & ANDERSEN, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt test. *J. consult. Psychol.*, 1950, 14, 197–198.

125. HARROWER-ERICKSON, M. R. Personality changes accompanying cerebral lesions: II. Rorschach studies of patients with focal epilepsy. *Arch. Neurol. Psychiat.*, 1940, 43, 1081–1107.

126. HARROWER-ERICKSON, M. R. Personality changes accompanying organic brain lesions: III. A study of preadolescent children. *J. genet. Psychol.*, 1941, 58, 391–405.

127. HARROWER, M. R. The results of psychometric and personality tests in multiple sclerosis. *Res. Publ. Ass. nerv. ment. Dis.*, 1950, 28, 461–470.

128. HEAD, H. Aphasia and kindred disorders of speech. Cambridge: Cambridge Univer. Press, 1926.

129. HEATH, R. G., & POOL, J. L. Bilateral frontal resection of frontal cortex for the treatment of psychoses. *J. nerv. ment. Dis.*, 1948, 107, 411–429.

130. HEATH, S. R. Clinical significance of motor defect, with military implications. *Amer. J. Psychol.*, 1944, 57, 482–499.

131. HEATH, S. R. A mental pattern found in motor deviates. *J. abnorm. soc. Psychol.*, 1946, 41, 223–225.

132. HEBB, D. O. Intelligence in man after large removals of cerebral tissue: report of four left frontal lobe cases. *J.*

gen. Psychol., 1939, 21, 73–87.

133. HEBB, D. O. The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proc. Amer. phil. Soc.*, 1942, 85, 275–292.

134. HEBB, D. O. Man's frontal lobes. *Arch. Neurol. Psychiat.*, 1945, 54, 10–24.

135. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

136. HEBB, D. O. Comparative and physiological psychology. *Annu. Rev. Psychol.*, 1950, 1, 173–188.

137. HEIDBREDER, EDNA. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, 52, 1–22.

138. HEWSON, L. The Wechsler-Bellevue scale and the substitution test as aids in neuro-psychiatric diagnosis. *J. nerv. ment. Dis.*, 1949, 109, 158–183, 246–265.

139. HOEDEMAKER, E., & MURRAY, M. Psychological tests in the diagnosis of organic brain disease. *Neurology*, 1952, 2, 144–153.

140. HOLBOURN, A. H. S. Mechanics of head injuries. *Lancet*, 1943, 2, 438–441.

141. HOWSON, J. D. Intellectual impairment associated with brain-injured patients as revealed in the Shaw Test of abstract thought. *J. Psychol.*, 1948, 2, 125–133.

142. HOYT, R., ELLIOT, H., & HEBB, D. O. The intelligence of schizophrenic patients following lobotomy treatment. *Queen Mary Vet. Hosp. Serv. Bull.*, *Montreal*, 1951, 6, 553–557.

143. HUGHES, R. M. Rorschach signs for the diagnosis of organic pathology. *Rorschach Res. Exch.*, 1948, 12, 165–167.

144. HUNT, H. F. A note on the problem of brain damage in rehabilitation and personnel work. *J. appl. Psychol.*, 1945, 29, 282–288.

145. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1032.

146. HUTTON, E. L. Results of prefrontal leucotomy. *Lancet*, 1943, 1, 362–366.

147. HUTTON, E. L., & BASSETT, M. Effect of leucotomy on creative personality. *J. ment. Sci.*, 1948, 94, 332–350.

148. INGHAM, S. D. Head injuries in relation to psychoneurotic symptoms and personality changes. *Bull. Los Angeles neurol. Soc.*, 1944, 9, 61–64.

149. JACKSON, J. H. *Selected writings of J. Hughlings Jackson.* London: Hodder & Stoughton, 1931–1932.

150. JAFFE, R. Influence of cerebral trauma on kinesthetic after-effects. *Amer. Psychologist*, 1951, 7, 265 (Abstract).

151. JEFFRESS, L. A. (Ed.) *Cerebral mechanisms in behavior. The Hixon Symposium.* New York: Wiley, 1951.

152. JONES, H. E., & CONRAD, H. S. The growth and decline of intelligence *Genet. Psychol. Monogr.*, 1933, 13, 223–298.

153. JONES, R. E. Personality changes in psychotics following prefrontal lobotomy. *J. abnorm. soc. Psychol.*, 1949, 44, 315–328.

154. JUCKEM, H., & WOLD, J. A. A study of the Hunt-Minnesota Test for organic brain damage at the upper levels of vocabulary. *J. Psychol.*, 1948, 12, 53–57.

155. KARNOSH, L. J., & GARDNER, W. J. An evaluation of the physical and mental capabilities following removal of the right cerebral hemisphere. *Cleveland clin. Quart.*, 1941, 8, 94–106.

156. KASS, W. Wechsler's Mental Deterioration Index in the diagnosis of organic brain disease. *Trans. Kansas Acad. Sci.*, 1949, 52, 66–70.

157. KENDALL, B. S., & GRAHAM, F. K. Further standardization of memory-for-designs test. *J. consult. Psychol.*, 1948, 12, 349–354.

158. KING, H. E., CLAUSEN, J., & SCARFF, J. F. Cutaneous thresholds for pain before and after unilateral prefrontal lobotomy: a preliminary report. *J. nerv. ment. Dis.*, 1950, 112, 93–96.

159. KISKER, G. W. Remarks on the problem of psychosurgery. *Amer. J. Psychiat.*, 1943, 100, 180–184.

160. KISKER, G. W. Abstract and categorical behavior following therapeutic brain surgery. *Psychosom. Med.*, 1944, 6, 146–150.

161. KISKER, G. W. The Rorschach analysis of psychotics subjected to neurosurgical interruption of the thalamo-cortical projections. *Psychiat. Quart.*, 1944, 18, 43–52.

162. KISKER, G. W. The behavioral sequelae of neurosurgical therapy: bilateral prefrontal lobotomy. *J. gen. Psychol.*, 1945, 33, 171–192.

163. KLAPPER, ZELDA S., & WERNER, H. Developmental deviations in brain-injured members of pairs of identical twins. *Quart. J. Child Behav.*, 1950, 2, 288–313.

164. KLEBANOFF, S. G. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.

165. KLEBANOFF, S. G. The Rorschach test in an analysis of personality changes in general paresis. *J. Pers.*, 1949, 17, 261–272.

166. KLEIN, G. S., & KRECH, D. "Cortical conductivity" in the brain-injured. *Amer. Psychologist*, 1951, **7**, 264. (Abstract)

167. KLEIN, R. Loss of written language due to dissolution of the phonetic structure of the word in brain abscess. *J. ment. Sci.*, 1951, 97, 328–339.

168. KOFF, S. A. The Rorschach test in the differential diagnosis of cerebral concussion and psychoneurosis. *Bull. U. S. Army med. Dep.*, 1946, 5, 170–173.

169. KOGAN, K. L. The personality reactions of children with epilepsy, with special reference to the Rorschach method. *Res. Publ. Ass. nerv. ment. Dis.*, 1947, 26, 616–630.

170. KÖHLER, W. Relational determination in perception. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951. Pp. 200–230.

171. KOLB, L. C. An evaluation of lobotomy and its potentialities for future research in psychiatry and the basic sciences. *J. nerv. ment. Dis.*, 1949, 110, 112–148.

172. KOSKOFF, Y. D., DENNIS, W., LAZOVIK, D., & WHEELER, E. T. The psychological effects of frontal lobotomy performed for alleviation of pain. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, 27, 723–753.

173. KOUNIN, J. S. Intellectual development and rigidity. In R. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child development and behavior*. New York: McGraw-Hill, 1943. Pp. 179–197.

174. KRECH, D. Dynamic systems as open neurological systems. *Psychol. Rev.*, 1950, 57, 345–361.

175. KROL, V. A., & DORKEN, H. The influence of subcortical (diencephalic) brain lesions on emotionality as reflected in the Rorschach color responses. *Amer. J. Psychiat.*, 1951 107, 839–843.

176. KROUT, M. H. Is the brain-injured a mental defective? *Amer. J. ment. Def.*, 1949, 54, 81–85.

177. LENOX, W. G. Seizure states. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 922–967.

178. LEVI, J., OPPENHEIM, S., & WECHSLER, D. Clinical use of the Mental Deterioration Index of the Bellevue-Wechsler scale. *J. abnorm. soc. Psychol.*, 1945, 40, 405–407.

179. LEVINE, K. N. A comparison of graphic Rorschach productions with scoring categories of the verbal Rorschach record in normal states, organic brain disease, neurotic and psychotic disorders. *Arch. Psychol.*, 1943, No. 282.

180. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

181. LEWINSKI, R. J. The psychometric pattern in epilepsy. *Amer. J. Orthopsychiat.*, 1947, 17, 714–722.

182. LIBERSON, W. T. Abnormal brain waves and intellectual impairment. *Inst. of Living*, 1944, No. 12, 234–248.

183. LIDZ, T., GAY, J. R., & TIETZE, C. Intelligence in cerebral deficit states and schizophrenia measured by Kohs Block test. *Arch. Neurol. Psychiat.*, 1942, **48**, 568–582.

184. LINDBERG, B. J. On the question of psychologically conditioned features in the Korsakow syndrome. *Acta psychiat., Kbh.*, 1946, 21, 497–542.

185. LISANSKY, E. S. Convulsive disorders and personality. *J. abnorm. soc. Psychol.*, 1948, **43**, 29–37.

186. LORD, E., & WOOD, L. Diagnostic values in a visuo-motor test. *Amer. J. Orthopsychiat.*, 1942, 12, 414–429.

187. LOVTSKAIA, A. J. The functions of the occipital lobe. *Neuropat. Psikhiat.*, 1944, 13, 78–80.

188. LUTZ, J. Zur psychischen Symptomatologie eines Schadebruches bei einem 1:1 alten Kinde. (Concerning the psychical symptomatology of skull fracture in a 13-month-old child.) *Z. Kinderpsychiat.*, 1949, 15, 173–185.

189. LUTZ, J. Psychische Symptome und Rekonvaleszenz nach Contusio Cerebri bei einem 6 Jahre alten Madelein. (Psychical symptoms and convalescence after cerebral concussion in a six-year-old girl.) *Z. Kinderpsychiat.*, 1949, 16, 97–109.

190. LYNN, J. G., LEVINE, K. N., & HEWSON, L. R. Psychologic tests for the clinical evaluation of the late "diffuse organic," "neurotic," and "normal" reactions after closed head injury. *Res. Publ. Ass. nerv. ment. Dis.*, 1945, 24, 296–378.

191. McCULLOUGH, M. W. Wechsler-Belle-

vue changes following prefrontal lobotomy. *J. clin. Psychol.*, 1950, **3**, 270–273.

192. McFIE, J., & PIERCY, M. F. The relation of laterality of lesion to performance on Weigl's sorting test. *J. ment. Sci.*, 1952, **98**, 299–305.

193. MACHOVER, KAREN. A case of frontal lobe injury following attempted suicide. *Rorschach Res. Exch.*, 1947, **11**, 9–20.

194. McKENZIE, K. G., & PROCTOR, L. D. Bilateral frontal leucotomy in the treatment of mental disease. *Canad. med. Ass. J.*, 1946, **55**, 433.

195. MAGARET, ANN. Parallels in the behavior of schizophrenics, paretics, and presenile non-psychotic patients. *J. abnorm. soc. Psychol.*, 1942, **37**, 511–528.

196. MALAMUD, RACHEL F. Validity of the Hunt-Minnesota Test for organic brain damage. *J. appl. Psychol.*, 1946, **30**, 271–275.

197. MALMO, R. B. Psychological aspects of frontal gyrectomy and frontal lobotomy in mental patients. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 537–564.

198. METARAZZO, J. D. A study of the diagnostic possibilities of the C. V. S. with a group of organic cases. *J. clin. Psychol.*, 1950, **6**, 337–343.

199. METTLER, F. A. (Ed.) *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.

200. METTLER, F. A. (Ed.) *Psychosurgical problems.* New York: Blakiston, 1952.

201. MEYER, A., & McLARDY, T. Leucotomy as an instrument of research. Neuropathological studies. *Proc. Roy. Soc. Med.*, 1947, **40**, 145.

202. MIXTER, W. J., TILLOTSON, K. J., & WEIS, D. Reports of partial frontal lobectomy and frontal lobotomy performed on three patients: one chronic epileptic and two cases of chronic depression. *Psychosom. Med.*, 1941, **3**, 26–37.

203. NAPOLI, P. J., & SWEENEY, L. Hostility in the chronic neurological patient. Paper read at East. Psychol. Ass., Atlantic City, 1952.

204. NATHANSON, M., & WORTIS, S. B. Severe rigidity in performance and thought in a case of presenile degenerative disease. *J. nerv. ment. Dis.*, 1948, **108**, 399–408.

205. NICHOLS, I. C., & HUNT, J. McV. A case

of partial bilateral frontal lobectomy. *Amer. J. Psychiat.*, 1940, **96**, 1063–1083.

206. OLTMAN, JANE E., BRODY, B. S., FRIEDMAN, S., & GREEN, W. F. Frontal lobotomy. *Amer. J. Psychiat.*, 1949, **105**, 742–751.

207. OSTRANDER, JESSIE M. Rorschach record from a patient after removal of a tumor from the frontal lobe. *Amer. Psychologist*, 1947, **2**, 406. (Abstract)

208. PARSONS, F. H. Modifications of design block performance before and after corpus callosum section. *Psychol. Bull.*, 1942, **39**, 494. (Abstract)

209. PARSONS, F. H. Eight cases of section of corpus callosum in individuals with a history of epileptic seizures: psychological tests. *J. gen. Psychol.*, 1943, **29**, 227–241.

210. PARTRIDGE, M. *Prefrontal leucotomy.* Springfield, Ill.: Charles C Thomas, 1950.

211. PATERSON, A., & ZANGWILL, O. L. Disorders of visual space perception associated with lesions of the right cerebral hemisphere. *Brain*, 1944, **67**, 331–358.

212. PENFIELD, W. Symposium on gyrectomy. Part I: Bilateral frontal gyrectomy and postoperative intelligence. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 519–534.

213. PETRIE, A. Personality changes after pre-frontal leucotomy. *Brit. J. med. Psychol.*, 1949, **22**, 200–207.

214. PETRIE, A. Preliminary report of changes after pre-frontal leucotomy. *J. ment. Sci.*, 1949, **95**, 449–455.

215. PEYTON, W. T., NORAN, H. H., & MILLER, E. W. Prefrontal lobectomy. *Amer. J. Psychiat.*, 1948, **104**, 513–523.

216. PFLUGFELDER, G. Intellektuelle Storungen nach schweren Schadeltraumen. *Mschr. f. Psychiat. Neurol.*, 1949–1950, **118–119**, 288–304, 378–404.

217. PINKERTON, P., & KELLY, J. An attempted correlation between clinical and psychometric findings in senile-arteriosclerotic dementia. *J. ment. Sci.*, 1952, **98**, 244–255.

218. PIOTROWSKI, Z. The Rorschach inkblot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.

219. PORTEUS, S. D. Thirty-five years experience with the Porteus Maze. *J. abnorm. soc. Psychol.*, 1950, **45**, 396–401.

220. PORTEUS, S. D., & KEPNER, R. DeM. Mental changes after bilateral pre-

frontal lobotomy. *Genet. Psychol. Monogr.*, 1944, **29**, 3–115.

221. PORTEUS, S. D., & PETERS, H. N. Maze test validation and psychosurgery *Genet. Psychol. Monogr.*, 1947, **36**.

222. PROBST, H. Über psychische Folgen des Schadelleruches im Kindesalter. *Z. Kinderpsychiat.*, 1949, **15**, 186–192.

223. RABIN, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.

224. RASHKIS, H. A. Three types of thinking disorder. *J. nerv. ment. Dis.*, 1947, **106** 650–670.

225. RASHKIS, H., CUSHMAN, J., & LANDIS, C. A new method for studying disorders of conceptual thinking. *J. abnorm. soc. Psychol.*, 1946, **41**, 70–82.

226. REINER, E. R., & SANDS, S. L. Lobotomy and psychopathology. *Arch. Neurol. Psychiat.*, 1951, **65**, 48–53.

227. REYNELL, W. R. A psychometric method of determining intellectual loss following head injury. *J. ment. Sci.*, 1944, **90**, 710–719.

228. ROBINSON, MARY F. What price lobotomy? *J. abnorm. soc. Psychol.*, 1946, **41**, 421–436.

229. ROGERS, L. S. A comparative evaluation of the Wechsler-Bellevue Mental Deterioration Index for various adult groups. *J. clin. Psychol.*, 1950, **6**, 199–202.

230. ROGERS, L. S. A note on Allen's index of deterioration. *J. clin. Psychol.*, 1950. 6. 203.

231. ROSS, W. O., & ROSS, S. Some Rorschach ratings of clinical value. *Rorschach Res. Exch.*, 1944, **8**, 1–9.

232. ROSVOLD, H. E., & MISHKIN, M. Evaluation of the effects of prefrontal lobotomy on intelligence. *Canad. J. Psychol.*, 1950, **4**, 122–126.

233. RUESCH, J., & BOWMAN, K. M. Prolonged post-traumatic syndromes following head injury. *Amer. J. Psychiat.*, 1945, **102**, 145–164.

234. RUESCH, J., & MOORE, B. E. Measurement of intellectual functions in the acute stage of head injury. *Arch. Neurol. Psychiat.*, 1943, **50**, 165–170.

235. RUST, R. M. Some correlates of the movement response. *J. Pers.*, 1948, **16**, 369–401.

236. RYLANDER, G. *Personality changes after operations on the frontal lobes.* London: Oxford, 1939.

237. RYLANDER, G. Mental changes after excision of cerebral tissue. A clinical study of 16 cases of resections in the parietal, temporal, and occipital lobes. *Acta Psychiat. Neurol.*, 1943, Suppl. 20.

238. RYLANDER, G. Psychological tests and personality analyses before and after frontal lobotomy. *Acta Psychiat. Neurol.*, 1947, Suppl. 47, 383–398.

239. SANDS, H., & PRICE, J. C. A pattern analysis of the Wechsler-Bellevue Adult Intelligence Scale in epilepsy. *Res. Publ. Assoc. nerv. ment. Dis.*, 1947, **26**, 604–615.

240. SCHEERER, M. Problems of performance analysis in the study of personality. *Ann. New York Acad. Sci.*, 1946, **46**, 653–678.

241. SCHEERER, M. An experiment in abstraction; testing form-disparity tolerance. *Confinia Neurol.*, 1949, **9**, 232–254.

242. SCHEERER, M., ROTHMANN, E., & GOLDSTEIN, K. A case of "idiot-savant." An experimental study of personality organization. *Psychol. Monogr.*, 1945, **58**, No. 4 (Whole No. 269).

243. SCHILDER, P. Congenital alexia and its relation to optic perception. *J. genet. Psychol.*, 1944, **65**, 67–88.

244. SCHRADER, P. J., & ROBINSON, M. F. An evaluation of prefrontal lobotomy through ward behavior. *J. abnorm. soc. Psychol.*, 1945, **40**, 61–69.

245. SHEPS, J. G. Intelligence of male non-institutionalized epileptics of military age. *J. ment. Sci.*, 1947, **93**, 82–88.

246. SHORVON, H. J. Prefrontal leucotomy and the depersonalization syndrome. *Lancet*, 1947, **253**, 714–718.

247. SILVER, A. A. Diagnosis and prognosis of behavior disorder associated with organic brain disease in children. *J. insurance Med.*, 1951, **6**, 38–42.

248. SINGER, J. L., MELTZOFF, J., & GOLDMAN, G. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359–364.

249. SLOAN, W., & BENSBERG, G. J. The stereognostic capacity of brain injured as compared with familial mental defectives. *J. clin. Psychol.*, 1951, **7**, 154–156.

250. SMITH, K. U. Bilateral integrative action of the cerebral cortex in man in verbal association and sensori-motor coordination. *J. exp. Psychol.*, 1947, **3**, 367–376.

251. SMITH, K. U., & AKELAITIS, A. J. Stud-

ies in the corpus callosum: I. Laterality in behavior and bilateral motor organization in man before and after section of the corpus callosum. *Arch. Neurol. Psychiat.*, 1942, **47**, 519–543.

252. SOMERFELD-ZISKIND, E., & ZISKIND, E. Effect of phenobarbital on the mentality of epileptic patients. *Arch. Neurol. Psychiat.*, 1940, **43**, 70–79.

253. SPIEGEL, E. Physiologic and psychologic results of thalamotomy. *Arch. Neurol. Psychiat.*, 1950, **64**, 306–307. (Abstract)

254. STENGEL, E. Loss of spatial orientation, constructional apraxia, and Gerstmann's syndrome. *J. ment. Sci.*, 1944, **90**, 753–760.

255. STRAUSS, A. A. Ways of thinking in brain-crippled deficient children. *Amer. J. Psychiat.*, 1944, **100**, 639–647.

256. STRAUSS, A. A., & LEHTINEN, L. Psychopathology and education of the brain-injured child. New York: Grune & Stratton, 1947.

257. STRAUSS, A. A., & WERNER, H. Deficiency in the finger schema in relation to arithmetic disability. *Amer. J. Orthopsychiat.*, 1938, **8**, 719–724.

258. STRAUSS, A. A., & WERNER, H. Finger agnosia in children. *Amer. J. Psychiat.*, 1939, **34**, 37–62.

259. STRAUSS, A. A., & WERNER, H. Disorders of conceptual thinking in the brain-injured child. *J. nerv. ment. Dis.*, 1942, **96**, 153–172.

260. STRAUSS, A. A., & WERNER, H. Comparative psychopathology of the brain-injured child and the traumatic brain-injured adult. *Amer. J. Psychiat.*, 1943, **99**, 835–838.

261. STRECKER, E. A., PALMER, H. D., & GRANT, F. C. Study of prefrontal lobotomy. *Amer. J. Psychiat.*, 1942, **98**, 524–532.

262. STROM-OLSEN, R., LOST, S. L., BRODY, M. B., & KNIGHT, G. C. Results of prefrontal leucotomy in 30 cases of mental disorder. *J. ment. Sci.*, 1943, **89**, 165–181.

263. TEUBER, H. L. Neuropsychology. In *Recent advances in diagnostic psychological testing: a critical summary*. Springfield, Ill.: Charles C Thomas, 1950.

264. TEUBER, H. L., BATTERSBY, W. S., & BENDER, M. B. Performance of complex visual tasks after cerebral lesions. *J. nerv. ment. Dis.*, 1951, **114**, 413–429.

265. TEUBER, H. L., & BENDER, M. B. The significance of changes in pattern vision following occipital lobe injury. *Amer. Psychologist*, 1946, **1**, 255. (Abstract)

266. TEUBER, H., & BENDER, M. B. Alterations in pattern vision following trauma of occipital lobes in man. *J. gen. Psychol.*, 1949, **40**, 35–57.

267. TEUBER, H. L., & BENDER, M. B. Neuro-ophthalmology: the oculomotor system. In E. A. Spiegel (Ed.), *Progress in neurology and psychiatry*. New York: Grune & Stratton, 1951. Pp. 148–177.

268. TEUBER, H. L., & BENDER, M. B. Performance of complex visual tasks after cerebral lesions. *Amer. Psychologist*, 1951, **7**, 265–266. (Abstract)

269. TRIST, E. L., TRIST, V., & BRODY, M. B. Discussion on the quality of mental test performance in intellectual deterioration. *Proc. Roy. Soc. Med.*, 1943, **36**, 243–252.

270. VAN DER LUGT, M. J. A. *The V. D. L. Psychomotor Scale*. Springfield, Mass.: Meed Scientific Apparatus Co., 1951.

271. VAN WATERS, R. O., & SACKS, J. G. Rorschach evaluation of the schizophrenic process following a prefrontal lobotomy. *J. Psychol.*, 1946, **25**, 73–88.

272. VIDOR, MARTHA. Personality changes following prefrontal leucotomy as reflected by the Minnesota Multiphasic Personality Inventory and the results of psychometric testing. *J. ment. Sci.*, 1951, **97**, 159–173.

273. WATTS, J. W., & FREEMAN, W. Intelligence following prefrontal lobotomy in obsessive tension states. *Arch. Neurol. Psychiat.*, 1945, **53**, 244–245.

274. WATTS, J. W., & FREEMAN, W. Psychosurgery for the relief of unbearable pain. *J. int. Coll. Surg.*, 1946, **9**, 679.

275. WECHSLER, D. *The measurement of adult intelligence*. Baltimore: Williams & Wilkins, 1944.

276. WECHSLER, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

277. WEIGL, E. On the psychology of the so-called processes of abstraction. *J. abnorm. soc. Psychol.*, 1941, **36**, 3–33.

278. WEINSTEIN, E. A., & KAHN, R. L. The syndrome of anosognosia. *Arch. Neurol. Psychiat.*, 1950, **64**, 772–791.

279. WEINSTEIN, E. A., & KAHN, R. L. Patterns of disorientation in organic brain disease. *J. Neuropath. clin. Neurol.*, 1951, **1**, 214–225.

280. WEINSTEIN, E. A., & KAHN, R. L. Non-aphasic misnaming (paraphasia) in organic brain disease. *Arch. Neurol. Psychiat.*, 1952, 67, 72–79.

281. WEISENBERG, T., & McBRIDE, KATHERINE. *Aphasia.* New York: Commonwealth Fund, 1935.

282. WERNER, H. Motion and motion perception: a study in vicarious functioning. *J. Psychol.*, 1945, 19, 317–327.

283. WERNER, H. Perceptual behavior of brain injured mentally deficient children. *Genet. Psychol. Monogr.*, 1945, 31.

284. WERNER, H. Abnormal and subnormal rigidity. *J. abnorm. soc. Psychol.*, 1946 41, 15–24.

285. WERNER, H. The concept of rigidity: a critical evaluation. *Psychol. Rev.*, 1946, 53, 43–52.

286. WERNER, H. *The comparative psychology of mental development.* Chicago: Follett, 1948.

287. WERNER, H. Thought disturbance with reference to figure-background impairment in brain-injured children. *Confina Neurol.*, 1949, 9, 255–263.

288. WERNER, H., & BOWERS, M. Auditory-motor organization in two clinical types of mentally deficient children. *J. genet. Psychol.*, 1941, 59, 85–99.

289. WERNER, H., & CARRISON, D. Measurement and development of the finger schema in mentally retarded children. *J. educ. Psychol.*, 1942, 33, 252–264.

290. WERNER, H., & CARRISON, D. Animistic thinking in brain-injured mentally retarded children. *J. abnorm. soc. Psychol.*, 1944, 39, 43–62.

291. WERNER, H., & STRAUSS, A. A. Problems and methods of functional analysis in mentally deficient children. *J. abnorm. soc. Psychol.*, 1939, 34, 37–62.

292. WERNER, H., & STRAUSS, A. A. Pathology of figure background relations in the child. *J. abnorm. soc. Psychol.*, 1941, 36, 236–248.

293. WERNER, H., & STRAUSS, A. A. Impairment in thought processes of brain-injured children. *Amer. J. ment. Def.*, 1943, 47, 291–295.

294. WERNER, H., & THUMA, B. D. A deficiency in the perception of apparent motion in children with brain injury. *Amer. J. Psychol.*, 1942, 55, 58–67.

295. WERNER, H., & THUMA, B. D. Critical flicker frequency in children with brain injury. *Amer. J. Psychol.*, 1942, 55, 394–399.

296. WERTHAM, F. The Mosaic Test: technique and psychopathological deductions. In L. E. Abt, & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 230–256.

297. WEXBERG, E. Testing methods for the differential diagnosis of mental deficiency in a case of arrested brain tumor. *Amer. J. ment. Def.*, 1941, 46, 39–45.

298. WINFIELD, D. Intellectual performance of cryptogenic epileptics, symptomatic epileptics, and post-traumatic encephalopaths. *J. abnorm. soc. Psychol.*, 1951, 46, 336–343.

299. WITTENBORN, J. R., & METTLER, F. A. Some psychological changes following psychosurgery. *J. abnorm. soc. Psychol.*, 1951, 46, 548–556.

300. WITTENBORN, J. R., BELL, E. G., & LESSER, G. S. Symptom patterns among organic patients of advanced age. *J. clin. Psychol.*, 1951, 7, 328–330.

301. WOLTMANN, A. G. The Bender Visual-Motor Gestalt Test. In L. E. Abt., & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1950. Pp. 322–356.

302. WORCHEL, P., & LYERLY, J. G. Effects of prefrontal lobotomy on depressed patients. *J. Neurophysiol.*, 1941, 4, 62–67.

303. WORTIS, S. B., BENDER, M. B., & TEUBER, H. L. The significance of extinction. *J. nerv. ment. Dis.*, 1948, 107, 382–387.

304. YACORZYNSKI, G. K., BOSHES, B., & DAVIS, L. Psychological changes produced by frontal lobotomy. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, 27, 642–657.

305. YACORZYNSKI, G. K., & DAVIS, L. Modification of perceptual responses in patients with unilateral lesions of the frontal lobes. *Psychol. Bull.*, 1942, 39, 493–494. (Abstract)

306. ZANGWILL, O. L. Observations on the Rorschach test in two cases of acute concussional head-injury. *J. ment. Sci.*, 1945, 91, 322–336.

307. ZANGWILL, O. L. A review of psychological work at the Brain Injuries Unit, Edinburgh, 1941–1945. *Brit. med. J.*, 1945, 2, 248–250.

INTELLECTUAL FUNCTION OF THE TEMPORAL LOBES[1]

BRENDA MILNER
McGill University

The clinical literature on the intellectual effects of human brain damage reveals a constant preoccupation with the problem of the role of the frontal lobes, with a corresponding neglect of other parts of the cerebral cortex. In particular, there is not a single systematic investigation of the effects of temporal-lobe damage in man, although there are several isolated and highly suggestive reports of individual cases. Fortunately the situation is quite different with regard to animal work, which in the last few years has yielded numerous reports dealing with the effects of temporal-lobe lesions of varying extent on the learning ability of lower primates. This material draws attention to types of deficit which might well be found at the human level also, but which have been neglected; for this reason the animal data will be presented in some detail, before passing to a review of the clinical literature.

Since this review deals only with cognitive functions, there will be no detailed discussion of the emotional changes often seen in temporal-lobe damage. Certain salient facts should, however, be noted. In the monkey, a decrease in emotional reactivity regularly follows deep-temporal removals (7, 17, 60, 80, 88, 105); in man, electrographic abnormality in the anterior temporal region is frequently associated with personality disturbances (4, 81), and ablation of the temporal poles may cause extreme docility (38). It is now believed that such phenomena depend largely on damage to the amygdala, as removal of this area in animals has striking, though inconsistent, emotional effects, causing savageness in the cat (8, 101) and tameness in the monkey in the studies mentioned above. The inconsistency is puzzling, but MacLean (74) points out that Schreiner, Kling, and Galambos have recently described increased docility in cats also after bitemporal excisions, although this does not appear in the published abstract (97). Stimulation of the amygdala in waking animals has also been carried out by Gastaut (35) for the cat, and by MacLean and Delgado (75) for both cat and monkey, and here there was no contradiction, well-organized rage responses being elicited from both species.

Such observations form part of broader inquiries into the role of the phylogenetically more primitive parts of the cerebrum in emotion. Thus Gastaut (35), in an admirable review of recent work on the rhinencephalon, points out the rich connections and structural complexity of the amygdala, which could well make this a key area in any emotional regulatory system. This literature has also been surveyed by MacLean (73, 74), but with theoretical speculations which are perhaps more elaborate than the data warrant. Further discussion would be out of place here.

42

Review of Animal Studies

The cognitive effects of temporal-lobe damage have been most clearly demonstrated in animal studies. A series of careful ablation experiments on the posterior cortex of the monkey has shown conclusively that the temporal lobes have important visual functions, and indeed that a major extrastriate focus for vision is to be found in this region. The experimental evidence also suggests that the temporal lobes play an appreciable, though minor, role in tactual discrimination.

Vision in Monkey after Removal of the Temporal Lobes

The first evidence that temporal-lobe lesions affect visual functions appears in the work of Sanger Brown and Schafer (17), who in 1888 reported gross, though transient, disturbance of visual object recognition in monkeys after removal of the temporal lobes. However, the importance of this finding does not seem to have been realized at the time, and no formal testing was done. Recent interest in the topic stems from Klüver and Bucy's (60) dramatic demonstration in 1938 that destruction of the temporal lobes in the macaque brings about a state of "psychic blindness," which they believed to be similar to that of visual agnosia in man. Their animals no longer discriminated between edible and inedible objects, nor between neutral and previously fear-provoking ones, and were markedly handicapped on tests of form discrimination; yet these disturbances could not be attributed to sensory loss, since the animals showed only slight upper-quadrant field defects and were abnormally reactive, rather than unresponsive, to visual stimuli.

One interpretation (33) of these results attributes the deficit to accidental damage to area 19, areas 18 and 19 being traditionally regarded as the visual association area, the seat of higher visual functions. However, the evidence is strongly against such an interpretation. Attempts to duplicate Klüver and Bucy's findings by extensive lesions within the prestriate cortex have failed. Large selective ablations of either the lateral or the mesial surfaces of areas 18 and 19 can be made with only slight visual disturbance (19, 64); and it has been claimed by Chow (21) that destruction of all but the inaccessible posterior ventromedial sector of the prestriate region only retards visual learning in so far as the animal may have to adjust to an excessively restricted visual field. As against these negative findings, the evidence for visual deficit following prestriate ablation is meager and internally inconsistent: Chow (19, 21) and Lashley (64) failed to confirm Ades's (1) report of amnesia for visual habits following one-stage bilateral prestriate ablation. Lashley's own finding of deficit on a visual conditional reaction has not been corroborated by Chow (21), and disturbance of object recognition in the first few postoperative days was seen by Chow but not by Lashley. According to Riopelle and Ades (unpublished study, cited in 92) there may be transient retardation of new learning, but there is no residual deficit. It is clear that the picture of mild, transient, and variable disturbance which emerges from these studies discredits the traditional view that the prestriate cortex is the major visual association area. Coupled with Klüver and Bucy's findings, it directs attention to the temporal lobes as a probable extrastriate focus of visual functions.

Klüver and Bucy's temporal-lobe extirpations were large, and included

the hippocampus and amygdaloid complex bilaterally, sparing only the primary auditory area (and most of the optic radiations). Later work has confirmed that such extensive damage to temporal neocortex and allocortex suffices to cause profound visual disturbance, even though care is taken not to invade the prestriate region (12, 80, 88, 90, 91). The disturbance has tended, however, to be less dramatic than that described by Klüver and Bucy, and only Mishkin (80) has reported persistent loss of object recognition, which is of course the essential feature of visual agnosia in man. In formal testing situations all animals show impairment on a variety of visual discrimination problems after removal of the temporal lobes; the deficit is apt to be particularly severe on tests of pattern discrimination but there may also be disturbance of color (12, 19, 80), and even of brightness discrimination (19, 80).

When attempting to evaluate the effects of temporal-lobe lesions, two points should be emphasized. First, the behavioral changes cannot be attributed to the slight and variable visual field defects that have been found, since much greater field defects may occur after parietal-cortex removal, but without producing such disturbance (12); the same is true in the case of extensive subtotal damage to the striate cortex (40, 59, 98). Secondly, the effects cannot be attributed to general intellectual deficit. The attempt to limit visual functions to the striate cortex, while ascribing visual deficits seen with extrastriate lesions to a nonspecific loss (65, 66), simply does not fit the facts. An animal may succeed on auditory or somesthetic problems at a time when it shows severe visual loss, and there is evidence that temporal-lobe damage need not prevent good performance on delayed-reaction problems (52, 80), although in some cases it apparently does (12). Thus we are forced to conclude that temporal-lobe ablation has a selectively detrimental effect on visual learning and retention.

More recent studies have sought to clarify the nature of this effect. Chow (20) has been able to show that the loss of previously learned visual discrimination habits after temporal-lobe removal is not due to the destruction of specific memory traces, since in one out of two monkeys further training on other discrimination problems reinstated the original habits. (The training is considered important, since no such dramatic recovery occurred in any of the three control animals.) But it is doubtful if anyone would defend the view that temporal-lobe ablation primarily destroys specific memory traces. On the contrary, most emphasis has been laid on the difficulty of new learning (60, 80, 90). In view of this, it is surprising to find Chow stating that there was no retardation of new learning, especially as one experimental animal completely failed to learn a new, or to relearn an old, pattern discrimination. It is true that the deficits he reports are in general mild, but this may well be due to the fact that his lesions spared the ventromedial portion of the temporal lobes, an area which, as we shall see later, is particularly important for visual functions. It should be added that the small number of animals and the marked individual differences in performance further limit the inferences which can usefully be drawn from this study.

Chow's own conclusion "that the temporal neocortex of monkey is not a storehouse for visual memories, but that it exerts some facilitative influence on the neural substrate for visual

discrimination" (20, p. 436) is cautious enough. Lashley (66), however, has gone a step further in citing Chow's data as evidence against the view that the temporal lobes have a visual function. The argument does not seem convincing. It rests on the assumption that if there is no loss of certain specific visual memories and no clear-cut sensory defect (such as a restricted visual field or a total loss of color vision), then we must be dealing with some quite general impairment, such as a loss of comparison attitude or a lowering of "vigilance." Such a view does scant justice to the complex nature of visual perception; it also ignores clear experimental evidence that temporal-lobe ablation does not affect all discrimination learning equally, but produces a marked deficit only on visual problems.

This deficit is admittedly hard to define. A visual disturbance which affects brightness, color, and pattern discrimination, while leaving object recognition and visual acuity intact, cuts across the classical distinction between complex "associative" functions and simpler sensory ones, and thus lends support to those who, on the basis of systematic studies of human brain damage, doubt the existence of pure disorders of visual integration (9, 10). It has been pointed out that we cannot decide intuitively what constitutes a simple, and what a complex visual task (19, 80). If complexity is measured in terms of difficulty for the intact animal, then Mishkin (80) finds, and Riopelle and Ades (90) agree, that, other things being equal, the degree of retardation on a series of visual discrimination problems after temporal-lobe ablation varies directly with the complexity of the problem. Yet we know that this is not so for all types of visual problem. Riopelle and

his associates (91), using patterned-string problems, found that the overall level of performance was not significantly changed by removal of the temporal lobes, but that the order of difficulty of the problems was not the same for the normal as for the operated monkeys. The data suggest that certain configurations which are perceptually misleading for the normal animal become less compelling after destruction of the temporal cortex.

The fact that the same operated monkeys failed to form learning sets when tested with a series of different object-discrimination problems is a new and interesting finding (91). Apparently the monkeys found the three hundredth problem as difficult as the first, although some progress was always made on the six trials allotted to any single problem. These observations are clearly important, but interpretation is complicated by the fact that the operated animals never reached the same level of proficiency as the normals on the individual problems, so that the interproblem transfer effects for the two groups cannot be compared fairly.

Visual Localization within the Temporal Lobe (Monkey)

The experiments just discussed were undertaken in order to define more precisely the nature of the deficit seen after destruction of the temporal lobes. The question next arises as to whether a comparably severe visual disturbance may be produced by smaller lesions within the massive temporal cortex, or whether the only important factor is the amount of temporal-lobe tissue destroyed.

Reviewing the literature with respect to the lateral surface, we find some inconsistencies. Klüver and Bucy (61) emphasize the lack of gross visual disturbance following bilateral removal of either the first, sec-

ond, or third temporal convolutions, and Ades and Raab (2) report that damage to the lateral surface had no effect on a learned pattern discrimination (they do not report the extent of the damage). Similarly, Mishkin (80) presents essentially negative results from lateral-surface extirpations, although he finds retardation in the mastery of new and difficult discrimination problems. However, Chow (19) describes somewhat more extensive impairment; after lesions in the region of the middle temporal convolution his monkeys showed deficits in retention and in subsequent reacquisition of various visual discrimination habits.

The effects of damage to the ventromedial surface of the temporal lobes are more consistent and more striking. The most systematic behavioral study is that of Mishkin (80) who reports that animals are as handicapped on formal tests of visual discrimination after ventral-surface extirpations as they are after complete removal of the temporal lobes, the only difference being that with the smaller lesions the visual disturbance is not apparent to casual inspection. Persistent impairment of form discrimination after damage to the "baso-medial" cortex of the temporal lobe is also reported by Poirier (88), thus confirming the importance of this area for visual functions.

A further problem of localization is raised by the suggestion that the peculiar effectiveness of deep temporal-lobe lesions may be due at least in part to destruction of the hippocampus (2, 60, 64). The hypothesis has been refuted by recent work. Mishkin (80) has data which indicate that where visual impairment follows removal of the hippocampus it is due to cutting of the white matter in the ventral surface of the temporal lobe, and not to destruction of the hippo-

campus per se. The fact that severing the fornix, and so interrupting the long projecting pathways of the hippocampus, has no effect on visually guided behavior (34) strengthens this interpretation.

The experimental evidence presented so far has established the special importance of the ventromedial part of the temporal lobe for visual discrimination; the lateral section is implicated to a lesser degree, and minimal disturbance follows damage to the prestriate cortex. But this account is incomplete, in that it is based on the effects of circumscribed lesions of temporal lobe or prestriate cortex. There is also evidence to show that posterior-cortex ablations which involve combined destruction of the prestriate and lateral-temporal neocortex are in fact exceedingly damaging to visual function.

Lashley and Clark (67) point out that there is no anatomical justification for subdividing the posterior association area of the macaque neocortex into separate parietal, temporal, and preoccipital regions. It is also known that only minor behavioral changes result from isolated destruction of any of these areas (2, 64, 94). On the basis of such facts Blum, Chow, and Pribram (12) argue that the posterior neocortex, exclusive of sensory projections, should be treated as a unit, and that, as such, it must be ablated in its entirety if we are to discover its true functional significance. They accordingly made extensive lesions on the lateral aspect of the posterior cortex, involving damage to parietal, temporal, and preoccipital regions, and thereby produced a very marked loss of discriminatory ability. It is remarkable that such lesions cause at least as much visual impairment as does complete removal of the temporal lobes, despite the fact that the important

ventromedial area of the temporal lobe is left intact. This finding has since been amply confirmed (92, 107); it is also consistent with an earlier report of permanent loss of a pattern-discrimination habit after combined destruction of the lateral-temporal and prestriate regions (2). It should be emphasized that these severe and persistent disturbances of discriminatory function do not automatically accompany any large cortical lesion, but are specific to bilateral posterior-cortex extirpations; only mild and essentially transient effects follow bilateral prefrontal or extensive unilateral ablations (92, 107). It is interesting that very similar conclusions have recently been reached concerning complex visual functions in man (104).

To recapitulate: it has been established that in the macaque visual learning ability depends on the integrity of a small area of extrastriate cortex, the ventromedial surface of the temporal lobe, while no small lesion elsewhere is crucial. Much larger destructions of lateral temporal surface and preoccipital cortex have equal or greater effect. It appears therefore that facilitation from the lateral-temporal-prestriate complex is required for the functioning of the ventromedial focus, or, as an alternative statement, that the two form a single system in which the smaller ventromedial area has a special importance.

Somesthetic Functions of the Temporal Lobe

While the ablation studies described above leave no doubt as to the importance of the temporal lobes for vision, the evidence is far less conclusive in the case of somesthesis. It is true that Klüver and Bucy (60) inferred from the behavior of their animals that they were unable to recognize objects tactually after re-moval of the temporal lobes, and thus that their "psychic blindness" was not exclusively visual. There was, however, no loss of cutaneous sensitivity, and unfortunately no complex tactual and kinesthetic tests were given. At first sight the suggestion that temporal-lobe damage may interfere with somesthetic functions is a little disconcerting, since the parietal lobe is the traditional locus of such functions, and the somesthetic importance of the posterior parietal lobule has long been recognized (94). Recently, however, Blum, Chow, and Pribram (12) and Blum (11) have shown that extension of parietal lesions into the posterior cortex of the temporal lobe causes a greater somesthetic impairment than do posterior parietal lesions alone; the loss was measured by means of an extensive test battery which included tactual latch-box and form-discrimination problems. These studies also show conclusively that in the monkey temporal-lobe ablation by itself is essentially without effect upon such tests; the most that could be demonstrated was a slight difficulty in the discrimination of roughness (12). It follows from these two findings that the temporal lobe is not focally concerned in tactual and kinesthetic functions, but that it can play a minor role, at least when the more important parietal area is excised.

Blum (11) has already commented on the similarity between these findings and those of Evans (28) on tactual and kinesthetic deficit in cases of human brain damage; both human and animal studies suggest that damage to cortex posterior and inferior to the acknowledged somesthetic area of the posterior parietal lobule increases the sensory deficit. There is also evidence that greater and more enduring somesthetic deficit results from cortical ablations in man than

in chimpanzee, and in chimpanzee than in monkey (11, 94). These facts combine to suggest that some somesthetic loss may be expected in man after removal of the temporal lobes, even without parietal injury.

THE PHYSIOLOGICAL PROBLEM

As has been seen, behavioral studies show beyond doubt that the temporal lobes mediate important visual functions; the question therefore arises as to the nature of the central nervous pathways involved. The first problem is to locate conduction pathways from the visual receiving area to the temporal lobes, which are able to survive extensive damage to areas 18 and 19. Data will be presented which argue strongly against an explanation in terms of transcortical systems alone, and which emphasize rather the role of the pulvinar in higher visual functions. A further problem is posed by the fact that similar visual deficit results from either small lesions on the ventral surface of the temporal lobes, or from much larger removals on the lateral surface of the posterior association cortex.

Strychnine neuronography has distinguished three zones within the temporal cortex of the macaque, and these differences appear to be correlated with differences in function. The three subdivisions are: the temporal pole; the supratemporal plane, together with the cortex on the superior lateral surface of the temporal lobe; and the main temporal sector comprising the ventral and posterolateral portions of the temporal lobe. It is this third section only which has been implicated in visual functions, and its organization is therefore of paramount interest here. What is known of the other two temporal regions may be briefly summarized. The temporal pole, an anatomically

distinct region, has connections with many areas, notably the orbitofrontal cortex and the amygdala (68, 89). It is chiefly concerned in autonomic and related somatic-motor functions (55, 75), though recent electrophysiological studies reveal diffuse connections between the temporal tip and the posterior association nuclei of the thalamus, suggesting that the polar region has more in common with the main temporal sector than is generally supposed (3). The second region, that of the supratemporal plane and adjacent cortex, contains the projection field of the medial geniculate, and subserves audition. Thorough studies have been made of its transcortical and intracortical connections (5, 102).

The main temporal region has many connections with other cortical areas, the most suggestive from the standpoint of visual function being those to and from the prestriate cortex. Such connections have been found for both the posterior lateral and the inferior temporal regions (6, 77, 87). In contrast to this it has generally been believed that there are few thalamic projections to this area (106), but such a view is no longer tenable. Both Chow (18) and Simpson (99) have provided anatomical evidence for projections from the posterior part of nucleus pulvinaris medialis to the lateral surface of the temporal lobe, and Chow has also confirmed earlier reports of projections from the posterior part of nucleus pulvinaris lateralis to the posterior temporal cortex. Furthermore, Pribram and Mishkin[2] have now found that, in the baboon, combined ventral surface and hippocampal lesions produce degeneration in the ventral half of nucleus pulvinaris inferior and also in the postero-

[2] Mishkin, M. Personal communication, 1953.

ventral half of nucleus pulvinaris lateralis. This is taken to indicate that Chow's original description of these projections, which he placed in the lateral temporo-occipital area, should be extended ventrally. We see, then, that both the inferior and the posterolateral portions of the temporal lobes are rich in transcortical and thalamocortical connections.

In trying to elucidate the role of this main temporal area in visual functions, we run at once into the following difficulty. It is known that the striate cortex (area 17) has no direct connections with the temporal lobes, but fires locally into area 18, which thus acts as a way-station for transcortical impulses from the primary receiving area (14, 22). Yet we have seen that extensive damage to areas 18 and 19 does not interfere greatly with visual learning; particularly striking is Evarts' (29) demonstration that extensive ablation of area 18 has no effect on either retention or postoperative learning of a "conditional" problem requiring the association of visual and auditory cues. An attempt to explain such data purely in terms of transcortical connections must assume that the small ventromedial section of area 18 remaining can function equipotentially for the whole; this seems particularly unlikely when we consider that only short association fibers are found in the visual cortex (22), and that the region of area 17 immediately adjacent to the intact remnant of area 18 in Evarts' experiment represents the most peripheral part of the visual field (76). One must therefore look elsewhere for pathways to account for extrastriate visual functions.

It seems most likely that transmission via subcortical relay and association nuclei will provide the answer. In support of this, recent physiologi-cal studies by Jasper and his coworkers may be cited (53). Working with monkeys under local anesthesia and sodium pentobarbital sedation, Jasper, Ajmone-Marsan, and Stoll (53) found that local electrical stimulation of area 17 produced strong firing in a narrow zone of the lateral geniculate and simultaneous firing of the adjacent portion of nucleus pulvinaris lateralis. In fact, their results suggest that after-discharge in area 17 is conducted more readily over subcortical pathways than transcortically to the parastriate area. As there are projections from nucleus pulvinaris lateralis to the posterior temporal region, it seems that we have here a means of by-passing the prestriate cortex. Arguing against this view, Chow (21) points out that prestriate ablations may cause degeneration in the lateral part of the pulvinar without there being any lasting visual deficit. However, in the one case of this kind reported, the locus of cell degeneration in the pulvinar was not coextensive with the area to which the striate cortex appears to fire (53), although admittedly there was some overlap. The evidence at present is not sufficient to settle the question finally one way or the other, but the apparent absence of adequate cortical connections in such cases is presumptive evidence for a subcortical relay of the type described. The fact that combined destruction of posterior-temporal and prestriate cortex causes permanent visual impairment supports this view, suggesting as it does that in such cases both the transcortical and the subcortical pathways from area 17 have been cut.

It remains to discuss briefly the second problem posed by behavioral studies: how destruction of a small area on the ventromedial surface of the temporal lobes, or interruption

of some of the pathways serving it, causes deficits as severe as those seen with extensive damage to the lateral surface of the posterior association cortex.   The behavioral evidence clearly shows that the ventromedial portion of the temporal lobe contains an important center for visual elaboration.   Physiological studies have revealed an area on the ventral surface to which impulses from the surrounding temporal cortex and from the prestriate area appear to converge, and which itself sends efferent impulses to the prestriate cortex (87). Ventromedial-surface extirpations could therefore be expected to interrupt many cortical circuits; it is probable that such lesions also disrupt circuits between temporal lobe and pulvinar.   Blum, Chow, and Pribram (12) and Chow (21) have denied the importance of these subcortical connections, but the evidence does not entirely justify their position.

The principal evidence cited by Blum, Chow, and Pribram is that there is no correlation between visual deficit and either extent or locus of degeneration in the pulvinar.   However, absence of retrograde degeneration in a thalamic nucleus is no final proof of lack of connections between that nucleus and the area of cortex destroyed. As Jasper, Ajmone-Marsan, and Stoll (53) have pointed out, projecting thalamocortical fibers may have widely distributed collaterals which may protect the cells of origin from degeneration.   This would explain the negative results of degeneration studies following moderate-sized ablations in the posterior association cortex. It is known that hemidecortication causes complete degeneration in the pulvinar, while ablations of parietal or temporal cortex alone have only minor effects (23, 24, 106). A further point is that loss of func-

tion may also be caused by interruption of corticofugal connections from temporal lobe to pulvinar.   Such corticofugal connections have been suggested by Jasper, Ajmone-Marsan, and Stoll in the study described above.

The main conclusions from this discussion can be summarized as follows. There is a focus for visual elaboration in the ventromedial section of the temporal lobe, and it is reasonable to believe that lesions in this area disrupt important corticocortical and thalamocortical circuits.   It has been argued that the deficit seen with extensive lateral surface ablations is partly due to the severing of pathways from area 17 to the temporal lobes; but there is no inconsistency in supposing that the lateral-temporal and prestriate areas are themselves implicated to some extent in the synthesis of visual perceptions.

REVIEW OF HUMAN CLINICAL STUDIES

In contrast to our present extensive knowledge of the deficits caused by temporal-lobe damage in monkeys, until recently we could only speculate as to the effects of such lesions in man. This ignorance was due in part to the clinical preoccupation with the frontal lobes, in part to a lack of adequate psychological testing when the temporal-lobe case was seen, and perhaps most of all to the unsatisfactory nature of the clinical material. The few psychological studies reported deal with cases of brain tumor, where it is notoriously difficult to assess the full extent of brain damage, and where increased intracranial pressure may produce disturbance far from the site of the primary lesion. Even in cases of penetrating head injury or of vascular accident, it is unlikely that damage will be neatly restricted to the temporal cortex, and,

failing autopsy, we can have only an approximate idea of the area of brain destroyed. The fact that bilateral temporal-lobe lesions in man are rare presents another obstacle to our understanding of temporal-lobe function, although the manifest nonequivalence of the two hemispheres, at least for language, suggests that unilateral studies may be more rewarding at the human than at the animal level.

Despite a lack of precise information, the gross symptomatology of temporal-lobe disease has long been known (32, 57, 62). The personality changes which are a conspicuous feature of human temporal-lobe pathology have already been mentioned. Sensory disturbances and complex visual hallucinations are equally striking symptoms and gain in significance when we consider that visual impairment is the characteristic feature of temporal-lobe damage in monkeys. Finally, there is aphasia, which often accompanies lesions of the left temporal lobe, and which will require special discussion later.

Among the sensory disturbances seen in cases of temporal-lobe tumor, auditory anomalies such as tinnitus are the most common (32), presumably providing the most reliable localizing sign. The occurrence of olfactory hallucinations is said to imply involvement of the uncus and adjoining rhinencephalic structures at the base of the temporal lobe. Homonymous visual field defects are another frequent symptom and Cushing states that these are usually limited to the upper quadrant contralateral to the lesion and typically show no macular sparing; such defects are taken to indicate damage to the optic radiations as they course round the inferior part of the lateral ventricle into the white matter of the temporal lobe (26). In agreement with this, Penfield and Rasmussen (86) find that cutting of the white matter below the cortex and five centimeters posterior to the temporal tip produces a homonymous hemianopsia; however, in these cases there appears to be sparing of the contralateral half of macular vision.

As for the more complex visual disturbances, it has been known for some time that various phenomena such as deja vu, micropsia and macropsia, and formed visual hallucinations, may be among the earliest signs of temporal-lobe disease (32, 42), and indeed this finding led Hauptmann (42) to suggest that the temporal region might mediate important visual functions. Moreover, in epileptic patients paroxysmal discharge within the temporal lobe may give rise to perceptual illusions or hallucinations; such states have been called "dream states" by Hughlings Jackson and "psychical seizures" by Penfield. It is interesting that in such patients electrical stimulation of the exposed temporal cortex may produce complex hallucinations, both auditory and visual, and sometimes appears to reactivate old memories. However, such effects are not elicited by stimulation of the temporal cortex in patients whose epileptic foci are elsewhere (86).

It is largely on the basis of these stimulation studies that Penfield has come to attribute to the temporal lobes a unique role in the preservation of memory traces, but the situation is far from simple. Certainly temporal-lobe removal in monkeys does not cause a general memory loss, but has a more specifically visual effect.

It is now time to consider what has been explicitly stated about intellectual functions in temporal-lobe disease. As was suggested earlier, the clinical literature of brain tumor

deals mainly with mental disturbance gross enough to disrupt everyday behavior. Summarizing the findings in such cases, Klebanoff (58) states that the presenting symptoms of intellectual deterioration are similar to, but on the whole less severe than, those seen in frontal-lobe tumor; thus the patient may appear confused, and may show memory disturbance and inability to concentrate. The fact that the onset of these nonspecific intellectual defects may be delayed in temporal-lobe cases has led to the view that they may be secondary effects due to involvement of the frontal lobes, an interpretation which clearly reflects the traditional view that the frontal lobes are the site of man's intellectual functions. There is indeed a suggestion that in the temporal-lobe cases the memory loss is apt to be particularly severe, affecting both recent and remote events (62), but the fact that removal of the tumor by resection of the temporal lobe normally causes an abatement of the gross symptoms of mental impairment advises caution in drawing such a conclusion.

In two studies where systematic mental testing was done following excision of temporal-lobe tumors, the aim was to study the function of the frontal rather than the temporal lobes. Thus Rylander (95) and Halstead (39) include temporal-lobe cases in their brain-damaged control groups, and both writers affirm that lesions outside the frontal lobes cause no impairment comparable with that found in their frontal-lobe cases. Yet the value of such assertions may be questioned. In the first place, Rylander deliberately excluded from his control group patients with damage to either the left-parietal or left-temporal lobes, on the grounds that in such cases surgical intervention

might greatly injure "the tools for mental activity." This treatment of language as a mere tool is dubious enough, but the risks inherent in such a restricted choice of subjects are here heightened by the use of a test battery composed largely of verbal tests. If, as will be suggested later, the right temporal lobe contributes mainly to the nonverbal aspects of intelligence, then Rylander has not adequately explored the probable areas of intellectual deficit in his seven patients with right temporal-lobe lesions. In Halstead's study, nine temporal-lobe cases are reported, but in only four of them (two left-sided and two right-) did the removals approach lobectomy. Of these, the two left-sided cases showed intellectual impairment at least equal to the average for the frontal-lobe group, while the right-sided cases gave negative results. This suggests that Halstead's special tests are in fact sensitive to left temporal-lobe destruction.

From this point onward the effects of right and left temporal-lobe damage will be treated separately: first, because of the prevalence of language disturbances (aphasia) after lesions of the left temporal lobe; and second, because of the possibility, hitherto unproven, that right temporal-lobe lesions may be more apt than left to cause deficit on certain nonlanguage tests.

Aphasia and the Left Temporal Lobe

Precise information concerning the location of a speech area within the temporal cortex comes from electrical stimulation of the exposed brain in epileptic patients. This technique has revealed an area in the posterior temporal lobe where stimulation may either arrest speech or cause the patient to become aphasic (86). The anterior limit of this region is ap-

proximately seven centimeters from the temporal tip, and excision of all cortex anterior to it can safely be made without causing more than a transient aphasia.

The literature of aphasia provides striking evidence that clean surgical destruction is less damaging than an active disease process. Thus Roberts (93) points out that the speech area in the posterior temporal cortex can be partially destroyed during the excision of an epileptogenic focus without any permanent aphasia, provided, that is, that the patient's seizures are stopped. Similarly, Fox and German (31) and Neilson and Raney (82) comment with surprise on the comparative mildness of the residual aphasia seen following complete resection of the left temporal lobe in cases of cerebral tumor, a finding which apparently contrasts sharply with the crippling effect upon language of even minor vascular lesions within the same area (82). The data are interpreted by these various authors as showing that the presence of partial damage within a speech area causes greater language disturbance than does the surgical removal of a much larger area. This is a special instance of a principle which has application beyond the field of language disorders: namely, that greater deterioration may result from the presence of abnormally functioning tissue than from mere absence of tissue (13, 45, 54).

If we now ask what the special characteristics of temporal-lobe aphasia are, we get no one clear-cut answer. Roberts (93) carried out a detailed study of 71 cases of aphasia (following excision of cortex adjoining speech areas in frontal, parietal, and temporal regions), and found no relationship between either the severity or the duration of the language disturbance and the site of the lesion.

This conclusion provides support for those who, like Goldstein (36, 37), have been consistently opposed to the doctrine that different language functions are differently localized in the left cerebral cortex. However, a more traditional view is that the aphasia seen in posterior lesions is primarily "receptive," (a disturbance of comprehension), as contrasted with the predominantly "expressive" disturbance seen with anterior lesions: Weisenberg and McBride (108) explicitly accept this formulation, although they are careful to point out that there is always some expressive defect, making the difference one of degree only. Head (43), who lays little stress on receptive disturbances, attributes his syntactic aphasia to lesions of the temporal lobe: a more popular view associates the temporal region with the auditory components of speech, with difficulty both in understanding and in retention of spoken language (31, 38, 56, 83, 96, 103). These studies as a whole permit the conclusion that in temporal-lobe aphasia speech comprehension is rather obviously impaired, but not in isolation from other aspects of language (96, 103).

We come now to the important question of how far aphasia (and here I include temporal-lobe aphasia) can be regarded as a purely linguistic disturbance which leaves "intelligence" essentially unimpaired, and how far it implies a more general intellectual loss. There are three main schools of thought: according to one (of which Goldstein is the best-known representative), aphasia is merely a manifestation of a more fundamental intellectual disorder (15, 36, 37, 110, 111); according to another, there may be loss, which does not affect all aspects of intelligence equally, so that the patient will do very poorly on some tests and comparatively well

on others (**31**, **46**, **108**); or again, there is the third possibility that aphasia is a mere loss of symbols, by which it is implied that the aphasic patient will have difficulties of communication, but no intrinsic difficulty in problem solving (**41**, **50**, **56**, **70**). Those who adopt this last position state that large cortical lesions *also* impair intelligence and that for this very reason such cases are not suitable for studying the typical patterning of abilities in aphasia (**41**, **56**). It will be argued here that it is the second viewpoint for which there is the most empirical support.

It is clear that the testing ground of these rival theories must be the domain of nonverbal skills. Here Kennedy and Wolf (**56**) make the important point that many tests purporting to be nonverbal are in fact not so, since the score depends upon how well the patient has understood the verbal instructions; hence, unless it can be shown in a given instance that the patient has grasped what is required, the work is invalid. This means, in effect, that the only unequivocal results will be those cases in which the patient succeeds on the easy items of a test (showing that he has in fact grasped the instructions), only to fail later on more difficult ones (**46**). Unfortunately, not all tests used in studies of aphasia meet this criterion.

If the results in the clinical literature are taken at their face value, there is still ample proof that aphasia is compatible with normal performance on some nonverbal tests. Meyers (**78**) finds that aphasic patients perform as well as their matched normal controls on nonverbal tests of inductive reasoning of the multiple-choice kind. Similarly, von Kuenburg (**63**) has failed to find clear-cut differences between aphasics and normal control subjects on sorting

tests calling for various classifications of colors and objects. Such studies are valuable in underlining the difficulty that many normal persons have on such tasks, a fact which is too often ignored in clinical studies.

Nevertheless it would be idle to deny that aphasic patients often show impairment on nonverbal tests. Van Woerkom (**110**) demonstrated such deficits in individual cases many years ago, and later work has essentially confirmed this finding. As was suggested earlier, the aim here is to show that the deficit is characteristically mild. Weisenberg and McBride (**108**) report that their aphasic patients consistently made better scores on nonverbal tests and on arithmetic than on verbal tests, but that these scores were still distinctly lower than would have been predicted in the case of a normal person with the same educational history. A similar picture emerges from Hebb's (**46**) study of six cases of aphasia following surgical excisions in the dominant hemisphere, and here repeated testing showed that the recovery of verbal and nonverbal functions followed the same time course. Finally, in a report of special interest for this present study, Fox and German (**31**) present follow-up observations in a case of left temporal lobectomy for cerebral tumor: they report "average" performance on a series of nonverbal tests at a time when the patient continued to do poorly on verbal and mathematical tests; at this time (fifteen months after operation), the patient still became confused if he listened to speech for any length of time, although the more overt signs of aphasia had disappeared. These results are not inconsistent with the view that there had been some loss of nonverbal ability, since Fox and German consider that their patient was originally above

average in intelligence, and thus assume that there had been some loss even in nonlanguage fields. It is of course evident that a major difficulty in the way of an accurate appraisal of all such data is the lack of any objective measure of the patients' intellectual status before the onset of aphasia.

It seems that the above studies, incomplete as they are, justify the view that in aphasia the pattern of test performance is likely to be one of severe deficit on verbal tests and relatively mild deficit on nonverbal ones. There is the further point that the degree of this disparity between verbal and nonverbal scores varies markedly from person to person, and with the tests used. These differences, which are very striking, are almost certainly related to the site of the lesion, but how, and to what extent, the data do not permit us to judge. An additional factor may be the degree to which the individual normally relies on verbal cues.

The studies outlined above show the typical patterning of test performance in cases of diffuse pathological damage to the left hemisphere, often involving the temporal cortex. It is reasonable to suppose that a clean surgical removal of the left temporal lobe (without brain tumor) would also affect verbal more than nonverbal performance, but that the over-all deficit would be less severe.

Special Functions of the Right Hemisphere

The absence of aphasia following lesions in the right hemisphere has led to a neglect of possible intellectual loss in such cases, and has probably encouraged the view that the right cerebral cortex plays an essentially ancillary role in intellectual functioning. It will be shown, on the contrary, that the evidence suggests

that damage to the posterior association cortex on the right produces deficit different in kind from that caused by comparable damage on the left, rather than an attenuated form of the same disturbance. The data relating specifically to the right temporal lobe are relatively meager, but they are completely consistent with the view just outlined.

The first suggestion that right hemisphere damage might cause its own characteristic deficit comes from Weisenberg and McBride's (108) detailed psychometric study of 22 cases of right-sided lesion without aphasia. Although these patients were slightly inferior to matched normal controls in occupational and educational level, this fact cannot account for the irregular nature of the test profiles obtained. Thus, while they approach the normals most closely on some verbal tests, they are significantly inferior on the arithmetic and most of the nonlanguage tests, notably the Porteus Maze Test. The authors conclude that the performance-test battery revealed a significant deficit in the appreciation and manipulation of forms and spatial relationships. These results are quite different from those obtained by the same workers with their aphasic group; the latter, it will be recalled, approached the normal most closely in nonlanguage work and arithmetic. We have no precise knowledge as to the extent of brain damage in the two groups, but Weisenberg and McBride explicitly state that they were ideally matched in this respect. However, it is not clear from the way the data are presented that the absolute mean scores of the right-hemisphere group on nonlanguage tests were lower than those of the aphasic group. If they were not, then we cannot conclude that the right hemisphere is more important than the left for the non-

language tests of Weisenberg and McBride's battery. All that has been shown is that these tests are more sensitive than verbal ones to right-hemisphere damage.

Further evidence on this point is provided by Hebb (44), in the case of a patient who had had a complete right temporal lobectomy for the removal of epileptogenic scar tissue. This patient performed well on all verbal tests, with Stanford-Binet Vocabulary at Superior-Adult level and full-scale IQ of 113, but revealed a startling deficit on tests of form perception, both visual and tactual, a finding which is in marked agreement with the results of Weisenberg and McBride. Hebb also reports that his patient had difficulty in following conversations involving more than two or three people, so that he would often make inappropriate remarks. This is an interesting observation, suggesting as it does a type of impairment which neither a conventional test battery nor a clinical interview could be expected to reveal. This study is valuable in that it deals with the effects of a well-localized lesion of one temporal lobe (although it should be added that there had also been a very small cortical excision in the precentral gyrus). Moreover, as the tests were carried out eight years after the temporal lobectomy, it seems safe to assume that the deficits observed were permanent. This study is unfortunately incomplete, in that there are no preoperative data. Yet it is worth noting that McFie, Piercy, and Zangwill (72) have described a similar picture of superior verbal intelligence combined with marked loss of "visuo-constructive ability," in the case of a 36-year-old engineer with a glioma of the right temporal lobe. However, they appear to interpret their findings as implying concomitant damage to the parietal lobe. In this instance the patient's profession seems reason enough for regarding the deficit as a direct result of the temporal-lobe tumor.

With the exception of a recent study by the writer (79), this exhausts the evidence relating visuospatial abilities specifically to the right temporal lobe. There remain a number of studies stressing the importance of the right hemisphere, and more particularly of the right parieto-occipital cortex, for effective spatial organization. The symptoms said to be more common in right- than in left-hemisphere lesions include: disturbance of the coordinates of visual space (49, 69); poor performance on visuo-constructive tasks such as map-drawing and block design (27, 49, 71, 72); and difficulty in putting on one's clothes, "apraxia for dressing" (27, 48). Such findings are of course not in accordance with the traditional view, which states that damage to the right parieto-occipital cortex merely causes neglect of the left side of the body and of the left half of the visual field (16, 25). Further, many of the cases on which the above reports are based involve massive, infiltrating tumors, so that bilateral damage can often be suspected. Yet this objection does not dispose of all cases; the growing number of such reports and the diversity of their sources support the idea that there is in fact a significantly higher incidence of visuo-constructive disorders in right-hemisphere lesions than in left.

A Recent Series of Temporal-Lobe Cases

This review may be concluded by summarizing the results of a recent study by the writer (79) of 25 cases of temporal-lobe operation by Penfield at the Montreal Neurological Institute. The study included 13 right temporal cases and 12 left, together

with 13 frontal and parietal cases as a control group. The operations were for the relief of focal epilepsy caused by atrophic lesions of the cortex usually dating from birth or early infancy (85). All patients were given a battery of sixteen tests before operation and again at the time of their discharge from the hospital about three weeks later, using equivalent forms of the same test wherever possible. The operations were carried out under local anesthesia, and Penfield attempted to reproduce the patient's aura by electrical stimulation of the exposed cortex; the epileptogenic area was then excised. It is important to note that the greatest abnormality was usually found on the ventral and mesial surfaces of the temporal lobe, and that the removal of tissue often extended to include the uncus, amygdaloid nucleus, and hippocampus.

The experimental and control groups were well matched with respect to age, vocabulary, and occupational status, as were also the right and left temporal-lobe groups. The average age for all subjects was 28 years, with a range from 14 to 45. The mean score on Form I of the Stanford-Binet Vocabulary (1916) was 28, which is slightly above the value for normal adults quoted by Weisenberg, Roe, and McBride (109). These facts are mentioned to show that we are not dealing with a senile or a mentally retarded group.

Two main results emerged. The first and most consistent was inferiority of the temporal-lobe group on two visual tests (the McGill Picture Anomaly [47] and the Wechsler Picture Arrangement), both of which involve complex picture material dealing with everyday social situations. This deficit was present before operation and therefore cannot be attributed to visual field defects.

The second finding was the marked inferiority of the right temporal-lobe group on tests of spatial patterning, both visual and tactual. These differences were present before operation, but were intensified after removal of the temporal lobe, a significant further loss being observed in the right temporal group on one test of block design.

That there should be a visual deficit in human temporal-lobe damage is completely consistent with the animal studies reviewed earlier in this paper, the only difference being that in the monkey there must be a bilateral lesion before any visual deficit can be discovered. However, in man, as in monkey, the precise nature of the deficit is not clear. We are justified in concluding that it is specifically visual and not a general impairment of "social intelligence," since no defect was seen on verbal tests, such as the McGill Verbal Situation (47), which are also designed to measure some kind of social awareness. On the other hand it is remarkable that no visual tests, other than those mentioned, discriminated between the temporal-lobe patients and the frontal and parietal, although the test battery included the Wechsler Picture Completion, the Benton Visual Memory, the C.M.M. Picture Analogies, and a shortened form of the Raven Progressive Matrices. It seems that the deficit appears when attention has to be given to many aspects of a complex picture, or when different pictures have to be arranged in meaningful order on the basis of slight differences of detail. Visual discriminations of this kind are well practiced in our culture, but they may be difficult to acquire initially. It is clear that much experimental work is needed before we can identify the factors making these tests difficult for patients with temporal-lobe

lesions. The schematic nature of the drawings, the presence or absence of social anomalies, and the mode of presentation of the test may all prove to be important.

The poor performance of the right temporal-lobe group on the Halstead Tactual Formboard (39) and on two visual tests of block design provides support for those who contend that the right posterior association cortex is more important than the left for the effective organization of space (48, 49, 71, 72).

However, not all patients in the right temporal group showed this spatial disability; the most severe disturbance occurred in cases where the removal extended far back along the inferior surface of the temporal lobe. This may mean either that the posterior temporal region is more important than the anterior for the perception of spatial patterns, or that the important factor is the size of the lesion in the right posterior cortex.

SUMMARY

In both monkey and man temporal-lobe damage can cause emotional and intellectual changes. The present paper has dealt only with the latter.

The main deficit is a visual one. In the monkey, bilateral removal of the temporal lobes severely hampers visual learning and retention, and in man even a unilateral lesion can impair the understanding of complex pictorial material. This loss is not an agnosia in the strict sense, nor is it a simple forgetting of learned discriminations.

Selective ablation studies in the monkey have shown that there is a "focus" on the ventral surface of the temporal lobe, damage to which is almost as detrimental to visual function as is complete lobectomy. Similar effects follow interruption of the transcortical and subcortical connections of this region. In human epileptic patients, the temporal-lobe cases showing pronounced visual deficit are those with focal lesions involving the inferior and mesial aspects of the temporal lobe.

These observations naturally raise the question of the central nervous pathways implicating the temporal cortex in specifically visual functions. It has been argued here that the data do not easily lend themselves to analysis in terms of transcortical systems alone, and that the possible importance of connections between the pulvinar and the temporal cortex should not be overlooked.

Any evaluation of temporal-lobe function in man requires a further breakdown into left and right temporal-lobe lesions. On the left, language disturbances of varying severity commonly follow lesions which interfere with the normal functioning of the speech area in the posterior temporal region, though it is pointed out that a clean surgical removal is apt to produce far milder effects than the continued presence of abnormally functioning tissue. On the right, there may be deficits of space perception, visual and nonvisual. Such deficits have also been found for right parieto-occipital lesions, so that we are justified in regarding the right posterior association cortex as specially important for the perception of spatial patterns.

REFERENCES

1. ADES, H. W. Effect of extirpation of parastriate cortex on learned visual discrimination in monkeys. *J. Neuropath. exp. Neurol.*, 1946, 5, 60–65.
2. ADES, H. W., & RAAB, D. H. Effect of preoccipital and temporal decortication on learned visual discrimination in monkeys. *J. Neurophysiol.*, 1949, 12, 101–108.
3. AJMONE-MARSAN, C., & STOLL, J., JR.

Subcortical connections of the temporal pole in relation to temporal lobe seizures. *Arch. Neurol. Psychiat., Chicago*, 1951, 66, 669–686.

4. BAILEY, P., & GIBBS, F. A. The surgical treatment of psychomotor epilepsy. *J. Amer. med. Ass.*, 1951, 145, 365–370.

5. BAILEY, P., BONIN. G. V., GAROL., H. W., & McCULLOCH. W. S. The functional organization of the temporal lobe of monkey (*Macaca mulatta*), and chimpanzee (*Pan satyrus*). *J. Neurophysiol.*, 1943, 6, 121–128.

6. BAILEY, P., BONIN, G. V., GAROL, H. W., & McCULLOCH, W. S. Long association fibres in cerebral hemispheres of monkey and chimpanzee. *J. Neurophysiol.*, 1943, 6, 129–134.

7. BARD, P. Central nervous mechanisms for the expression of anger in animals. In M. L. Reymert (Ed.), *Feelings and emotions: the Mooseheart symposium.* New York: McGraw-Hill, 1950. Pp. 211–237.

8. BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, 27, 362–404.

9. BAY, E. Agnosie und Funktionswandel. Eine hirnpathologische Studie. *Monogr. Ges. geb. Neur. Psychiat.*, Berlin, 1950, 73, 1–94.

10. BENDER, M. B., & TEUBER, H. L. Psychopathology of vision. *Progr. Neurol. Psychiat.*, 1949, 4, 163–192.

11. BLUM, JOSEPHINE S. Cortical organization in somesthesis: effects of lesions in posterior associative cortex on somatosensory function in *Macaca mulatta*. *Comp. Psychol. Monogr.*, 1951, 20, 219–249.

12. BLUM, JOSEPHINE S., CHOW, K. L., & PRIBRAM, K. L. A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *J. comp. Neurol.*, 1950, 93, 53–100.

13. BLUM, R. A. The effect of bilateral removal of the prefrontal granular cortex on delayed response and emotionality in chimpanzee. *Amer. Psychologist*, 1948, 3, 237–238. (Abstract)

14. BONIN, G. V., GAROL, H. W., & McCULLOCH, W. S. The functional organization of the occipital lobe. *Biol. Sympos.*, 1942, 7, 165–192.

15. BOUMAN, L., & GRÜNBAUM, A. A. Experimentell-psychologische Untersuchungen zur Aphasie und Para-

phasie. *Z. ges. Neurol. Psychiat.*, 1925, 96, 481–538.

16. BRAIN, R. Visual disorientation with special reference to lesions of the right cerebral hemisphere. *Brain*, 1941, 64, 244–272.

17. BROWN, SANGER, & SCHAFER, E. A. An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Philos. Trans.*, 1888, 179B, 303–327.

18. CHOW, K. L. A retrograde cell degeneration study of the cortical projection field of the pulvinar in the monkey. *J. comp. Neurol.*, 1950, 93, 313–340.

19. CHOW, K. L. Effects of partial extirpations of the posterior association cortex on visually mediated behavior in monkeys. *Comp. Psychol. Monogr.*, 1951, 20, 187–217.

20. CHOW, K. L. Conditions influencing the recovery of visual discriminative habits in monkeys following temporal neocortical ablations. *J. comp. physiol. Psychol.*, 1952, 45, 430–437.

21. CHOW, K. L. Further studies on selective ablation of associative cortex in relation to visually mediated behavior. *J. comp. physiol. Psychol.*, 1952, 45, 109–118.

22. CLARK, W. E. LE G. Observations on the association fibre system of the visual cortex and the central representation of the retina. *J. Anat., London*, 1940–41, 75, 225–235.

23. CLARK, W. E. LE G., & BOGGON, R. H. The thalamic connections of the parietal and frontal lobes of the brain in the monkey. *Philos. Trans.*, 1935, 224B, 313–359.

24. CLARK, W. E. LE G., & NORTHFIELD, D. W. C. The cortical projection of the pulvinar in the macaque monkey. *Brain*, 1937, 60, 126–142.

25. CRITCHLEY, M. In Discussion on parietal lobe syndromes. *Proc. roy. Soc. Med.*, 1951, 44, 337–346.

26. CUSHING, H. Distortions of the visual fields in cases of brain tumor. VI. The field defects produced by temporal lobe lesions. *Brain*, 1922, 44, 341–396.

27. DIDE, M. Les désorientations temporo-spatiales et la prépondérance de l'hémisphère droit dans les agnoso-akinésies proprioceptives. *Encéphale*, 1938, 33, 276–294.

28. EVANS, J. P. A study of the sensory defects resulting from the excision of cerebral substance in humans. *Res. Publ. Ass. nerv. ment. Dis.*, 1935, 15, 331–370.

29. Evarts, E. V. Effect of ablation of prestriate cortex on auditory-visual association in monkey. *J. Neurophysiol.*, 1952, 15, 191–200.

30. Forgays, D. G. Reversible disturbance of function in man following cortical insult. *J. comp. physiol. Psychol.*, 1952, 45, 209–215.

31. Fox, J. C., & German, W. J. Observations following left (dominant) temporal lobectomy. *Arch. Neurol. Psychiat., Chicago*, 1935, 33, 791–806.

32. Frazier, J. C., & Rowe, S. N. Localization of fifty-one verified tumors of the temporal lobe. *Res. Publ. Ass. nerv. ment. Dis.*, 1934, 13, 251–258.

33. Fulton, J. F. *Physiology of the nervous system.* (2nd Ed.) New York: Oxford Univer. Press, 1943.

34. Garcia-Bengochea, F., Corrigan, R., Morgane, P., Russell, D., & Heath, R. G. Studies on the function of the temporal lobes. I. The section of the fornix. *Trans. Amer. neurol. Ass.*, 1951, 76, 238–239.

35. Gastaut, H. Corrélations entre le système nerveux végétatif et le système de la vie de relation dans le rhinencéphale. *J. Physiol., Paris*, 1952, 44, 431–470.

36. Goldstein, K. Über Aphasie. *Schweiz. Neurol. Psychiat.*, 1926, 19, 3–38.

37. Goldstein, K. *Language and language disturbances.* New York: Grune and Stratton, 1948.

38. Green, J. R., Duisberg, R. E. H., & McGrath, W. B. Focal epilepsy of psychomotor type. A preliminary report of observations on effects of surgical therapy. *J. Neurosurg.*, 1951, 8, 157–172.

39. Halstead, W. C. *Brain and intelligence: A quantitative study of the frontal lobes.* Chicago: Univer. of Chicago Press, 1947.

40. Harlow, H. F. Recovery of pattern discrimination in monkeys following unilateral occipital lobectomy. *J. comp. Psychol.*, 1939, 27, 467–489.

41. Hauptmann, A. Ist die amnestische Aphasie Teilerscheinung einer Beeinträchtigung des "kategorialen" Verhaltens? *Mschr. Psychiat.*, 1931, 79, 302–313.

42. Hauptmann, A. Zur Symptomatologie der Erkrankungen des rechten Schläfenlappens. *Dtsch. Z. Nervenheilk.*, 1931, 117–119, 170–183.

43. Head, H. *Aphasia and kindred disorders of speech.* New York: Macmillan, 1926.

44. Hebb, D. O. Intelligence in man after large removals of cerebral tissue: defects following right temporal lobectomy. *J. gen. Psychol.*, 1939, 21, 437–446.

45. Hebb, D. O. Intelligence in man after large removals of cerebral tissue: report of four left frontal lobe cases. *J. gen. Psychol.*, 1939, 21, 73–87.

46. Hebb, D. O. The effect of early and late brain injury upon test scores and the nature of normal adult intelligence. *Proc. Amer. phil. Soc.*, 1942, 85, 275–292.

47. Hebb, D. O., & Morton, N. W. The McGill Adult Comprehension Examination: Verbal Situation and Picture Anomaly Series. *J. educ. Psychol.*, 1943, 34, 16–25.

48. Hécaen, H., & de Ajuriaguerra, J. L'apraxie de l'habillage. Ses rapports avec la planotopokinésie et les troubles de la somatognosie. *Encéphale*, 1942–1945, 8–10, 113–144.

49. Hécaen, H., de Ajuriaguerra, J., & Massonnet, J. Les troubles visuo-constructifs par lésion pariéto-occipitale droite. Rôle des perturbations vestibulaires. *Encéphale*, 1951, 40, 122–179.

50. Isserlin, M., Kuenburg, M.v., & Hofbauer. Zur Pathologie der Beziehungen zwischen Sprechen und Denken. *Zbl. ges Neurol. Psychiat.*, 1927, 47, 252–254.

51. Jackson, J. H. *Selected writings of John Hughlings Jackson.* London: Hodder and Stoughton, 1932.

52. Jacobsen, C. F., & Elder, J. H. Studies of cerebral function in primates: II. The effect of temporal lobe lesions on delayed response in monkeys. *Comp. Psychol. Monogr.*, 1936, 13, 61–65.

53. Jasper, H., Ajmone-Marsan, C., & Stoll, J. Corticofugal projections to the brain stem. *Arch. Neurol. Psychiat., Chicago*, 1952, 67, 155–177.

54. Jefferson, G. Removal of right or left frontal lobes in man. *Brit. med. J.*, 1937, 2, 199–206.

55. Kaada, B. R., Pribram, K. H., & Epstein, J. A. Vascular and respiratory effects of stimulation of the orbitofrontal and limbic areas. *J. Neurophysiol.*, 1949, 12, 347–356.

56. Kennedy, F., & Wolf, A. The relationship of intellect to speech defect in aphasic patients. *J. nerv. ment. Dis.*, 1936, 84, 125–145; 293–311.

57. Keschner, M., Bender, M. B., & Strauss, I. Mental symptoms in

cases of tumor of the temporal lobe. *Arch. Neurol. Psychiat., Chicago*, 1936, **35**, 572–593.

58. KLEBANOFF, S. G. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.

59. KLÜVER, H. Certain effects of lesions of the occipital lobes in macaques. *J. Psychol.*, 1937, **4**, 383–401.

60. KLÜVER, H., & BUCY, P. C. An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness." *J. Psychol.*, 1938, **5**, 33–54.

61. KLÜVER, H., & BUCY, P. C. Preliminary analysis of functions of the temporal lobe in monkeys. *Arch. Neurol. Psychiat., Chicago*, 1939, **42**, 979–1000.

62. KOLODNY, A. The symptomatology of tumors of the temporal lobe. *Brain*, 1928, **51**, 385–417.

63. KUENBURG, M. v. Zuordnungsversuche bei Gesunden und Sprachgestörten. *Arch. ges. Psychol.*, 1930, **76**, 257–352.

64. LASHLEY, K. S. The mechanism of vision: XVIII. Effects of destroying the visual "associative areas" of the monkey. *Genet. Psychol. Monogr.*, 1948, **37**, 107–166.

65. LASHLEY, K. S. In search of the engram. *Sympos. Soc. exp. Biol.*, 1950, **4**, 454–482.

66. LASHLEY, K. S. Functional interpretation of anatomic patterns. *Res. Publ. Ass. nerv. ment. Dis.*, 1952, **30**, 529–547.

67. LASHLEY, K. S., & CLARK, G. The cytoarchitecture of the cerebral cortex of Ateles: a critical examination of architectonic studies. *J. comp. Neurol.*, 1946, **85**, 223–305.

68. LENNOX, MARGARET A., DUNSMORE, R. H., EPSTEIN, J. A., & PRIBRAM, K. H. Electrocorticographic effects of stimulation of posterior orbital, temporal, and cingulate areas of *Macaca mulatta. J. Neurophysiol.*, 1950, **13**, 383–388.

69. LENZ, H. Raumsinnstörungen bei Hirnverletzungen. *Dtsch. Z. Nervenheilk.*, 1944, **157**, 22–64.

70. LOTMAR, F. Zur Pathophysiologie der erschwerten Wortfindung bei Aphasischen. *Schweiz. Arch. Neurol. Psychiat.*, 1933, **30**, 86–158; 322–379.

71. MCFIE, J., & PIERCY, M. F. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, **75**, 292–311.

72. MCFIE, J., PIERCY, M. F., & ZANGWILL, O. L. Visual-spatial agnosia associated with lesions of the right cerebral hemisphere *Brain*, 1950, **73**, 167–190.

73. MACLEAN, P. D. Psychosomatic disease and the "visceral brain." Recent developments bearing on the Papez theory of emotion *Psychosom. Med.*, 1949, **11**, 338–353.

74. MACLEAN, P. D. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain) *EEG clin. Neurophysiol.*, 1952, **4**, 407–418.

75. MACLEAN, P. D., & DELGADO, J. M. R. Electrical and chemical stimulation of frontotemporal portion of limbic system in the waking animal. *EEG clin. Neurophysiol.*, 1953, **5**, 91–100.

76. MARSHALL, W. H., & TALBOT, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. *Biol. Sympos.*, 1942, **7**, 117–164.

77. METTLER, F. A. Corticofugal fiber connections of the cortex of *Macaca mulatta*. The temporal region. *J. comp. Neurol.*, 1935, **63**, 25–47.

78. MEYERS, R. Relation of "thinking" and language: an experimental approach, using dysphasic patients. *Arch. Neurol. Psychiat., Chicago*, 1948, **60**, 119–139.

79. MILNER, BRENDA. Intellectual effects of temporal-lobe damage in man. Unpublished doctor's thesis, McGill Univer., 1952.

80. MISHKIN, M. Effects of selective ablations of the temporal lobes on the visually guided behavior of monkeys and baboons. Unpublished doctor's thesis, McGill Univer., 1951.

81. MULDER, D. W., & DALY, D. Psychiatric symptoms associated with lesions of the temporal lobe. *J. Amer. med. Ass.*, 1952, **150**, 173–176.

82. NEILSEN, J. M., & RANEY, R. B. Recovery from aphasia studied in cases of lobectomy. *Arch. Neurol. Psychiat., Chicago*, 1939, **42**, 189–200.

83. OBRADOR, S. Temporal lobotomy. *J. neuropath. exp. Neurol.*, 1947, **6**, 185–193.

84. PENFIELD, W. Discussion. In J. P. Evans, A study of the sensory defects resulting from the excision of cerebral substance in humans. *Res. Publ. Ass. nerv. ment. Dis.*, 1935, **15**, 366–370.

85. PENFIELD, W., & BALDWIN, M. Temporal lobe seizures and the technique of surgical therapy. *Ann. Surg.*, 1952, **136**, 625–634.

86. PENFIELD, W., & RASMUSSEN, T. *The cerebral cortex of man.* New York:

Macmillan, 1950.

87. Petr, R., Kralove, H., Holden, L., & Jirout, J. The efferent intercortical connections of the superficial cortex of the temporal lobe (*Macaca mulatta*). *J. neuropath. exp. Neurol.*, 1949, **8**, 100–103.

88. Poirier, L. J. Anatomical and experimental studies on the temporal lobes of the macaque. *J. comp. Neurol.*, 1952, **96**, 209–248.

89. Pribram, K. H., Lennox, Margaret A., & Dunsmore, R. H. Some connections of the orbito-fronto-temporal, limbic, and hippocampal areas of *Macaca mulatta*. *J. Neurophysiol.*, 1950, **13**, 127–136.

90. Riopelle, A. J., & Ades, H. W. Discrimination following deep temporal lesions. *Amer. Psychologist*, 1951, **6**, 261–262. (Abstract)

91. Riopelle, A. J., Alper, R. G., Strong, P. N., & Ades, H. W. Multiple discrimination and patterned string performance of normal and temporal-lobectomized monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 145–149.

92. Riopelle, A. J., Harlow, H. F., Settlage, P. H., & Ades, H. W. Performance of normal and operated monkeys on visual learning tests. *J. comp. physiol. Psychol.*, 1951, **44**, 283–289.

93. Roberts, L. Localization of speech in the cerebral cortex. *Trans. Amer. neurol. Ass.*, 1951, **76**, 43–50.

94. Ruch, T. C., Fulton, J. F., & German, W. J. Sensory discrimination in monkey, chimpanzee, and man after lesions of the parietal lobe. *Arch. Neurol. Psychiat.*, *Chicago*, 1938, **39**, 919–937.

95. Rylander, G. Mental changes after excision of cerebral tissue. A clinical study of 16 cases of resections in the parietal, temporal and occipital lobes. *Acta Psychiat. Neurol. Kbh.*, 1943, Suppl. 25, 1–81.

96. Schiller, F. Aphasia studied in patients with missile wounds. *J. Neurol. Neurosurg. Psychiat.*, 1947, **10**, 183–197.

97. Schreiner, L., Kling, A., & Galambos, R. Central nervous system lesions and aggressive behavior in cats. *Fed. Proc.*, 1952, **11**, 142. (Abstract)

98. Settlage, P. H. The effect of occipital lesions on visually guided behavior in the monkey. *J. comp. Psychol.*, 1939,

27, 93–131.

99. Simpson, D. A. The projection of the pulvinar to the temporal lobe. *J. Anat.*, *London*, 1952, **86**, 20–28.

100. Smith, W. K. Non-olfactory functions of the pyriform-amygdaloid-hippocampal complex. *Fed. Proc.*, 1950, **9**, 118. (Abstract)

101. Spiegel, E. A., Miller, H. R., & Oppenheimer, M. J. Forebrain and rage reactions. *J. Neurophysiol.*, 1940, **3**, 538–548.

102. Sugar, O., French, J. D., & Chusid, J. G. Corticocortical connections of the superior surface of the temporal operculum in the monkey (*Macaca mulatta*). *J. Neurophysiol.*, 1948, **11**, 175–192.

103. Teitelbaum, H. A. The principle of primary and associated disturbances of higher cortical functions as applied to temporal lobe lesions. *J. nerv. ment. Dis.*, 1942, **96**, 261–273.

104. Teuber, H. L., Battersby, W. S., & Bender, M. B. Performance of complex visual tasks after cerebral lesions. *J. nerv. ment. Dis.*, 1951, 114, 413–429.

105. Thomson, A. F., & Walker, A. E. Behavioral alterations following lesions of the medial surface of the temporal lobe. *Arch. Neurol. Psychiat.*, *Chicago*, 1951, **65**, 251–252.

106. Walker, A. E. *The primate thalamus.* Chicago: Univer. of Chicago Press, 1938.

107. Warren, J. M., & Harlow, H. F. Learned discrimination performance by monkeys after prolonged postoperative recovery from large cortical lesions. *J. comp. physiol. Psychol.*, 1952, **45**, 119–126.

108. Weisenberg, T., & McBride, Katharine E. *Aphasia: a clinical and psychological study.* New York: Commonwealth Fund, 1935.

109. Weisenberg, T., Roe, Anne, & McBride, Katharine E. *Adult intelligence: a psychological study of test performance.* New York: Commonwealth Fund, 1936.

110. Woerkom, W. v. Sur l'état psychique des aphasiques. *Encéphale*, 1923, **18**, 286–304.

111. Wolpert, I. Ueber das Wesen der literalen Alexie, Beitrag zur Aphasielehre. *Mschr. Psychiat.*, 1930, **75**, 207–266.

A GENERALIZATION OF SIDMAN'S RESULTS ON GROUP AND INDIVIDUAL FUNCTIONS, AND A CRITERION

DAVID BAKAN
University of Missouri

In a recent paper Sidman[1] has indicated a devastating criticism of a great deal of current and historical psychological research. He has demonstrated in the particular case of Hull's exponential growth function that if the functional relationship between two variables, x and y, for any particular individual is

$$y = f(x) \qquad [1]$$

then the average y-value, $\bar{y}$, is of the form

$$\bar{y} = g(x) \qquad [2]$$

where $g(x)$ may be a fundamentally different equation from $f(x)$. He shows that if the learning curve for an individual is

$$y = M - Me^{-kx} \qquad [3]$$

then the form of the function based on the average y is not necessarily the same as [3].

Sidman indicates that the kind of analysis he has made can be applied to any function. In this paper we attempt to generalize Sidman's results for any functional relationship, and to provide a criterion for deciding on the legitimacy of the averaging operation as a device for making inferences concerning individual functional relationships.

The major tool that we avail ourselves of is the Maclaurin series:

[1] Sidman, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, 49, 263–269.

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \cdots \qquad [4]$$

where $f(x)$ is our function, $f(0)$ is the value of the function when $x = 0$, $f'(0)$ is the value of the first derivative of the function when $x = 0$, etc.

Thus, if

$$y = f(a, b, c, \cdots, x) \qquad [5]$$

for a particular individual, then

$$\bar{y} = \frac{\sum y}{n} = \frac{\sum f(a, b, c, \cdots, x)}{n}. \qquad [6]$$

Applying Maclaurin's series, we get

$$y = y_0 + y_0'x + y_0''\frac{x^2}{2!} + y_0'''\frac{x^3}{3!} + \cdots \qquad [7]$$

and

$$\bar{y} = \bar{y}_0 + \bar{y}_0'x + \bar{y}_0''\frac{x^2}{2!} + \bar{y}_0'''\frac{x^3}{3!} + \cdots. \qquad [8]$$

The criterion which we offer to determine the legitimacy of the averaging operation as a basis for making inferences concerning individual functional relationships is as follows: If the coefficients $\bar{y}_0$, $\bar{y}_0'$, $\bar{y}_0''$, etc. are

simply functions of the average parameters, a, b, c, etc., then the operation may be considered legitimate. When this criterion is satisfied, the form of the average curve will be identical with the form of each individual curve. It will differ from the individual curves only in the parameters of the function. When the criterion is satisfied the parameter of the group function will be related to the parameter of any individual function in the same way that any mean is related to any individual score. If we restrict ourselves to functions which satisfy the criterion, we may then make inferences concerning the nature of individual functions from the nature of group functions, since the functions will have the same form.

To demonstrate the application of the criterion we will examine two functions

$$y = ax^2 + bx + c \qquad [9]$$

and

$$y = M - Me^{-kx}. \qquad [10]$$

If $y = ax^2 + bx + c$, then

$$y_0 = c$$
$$y' = 2ax + b \qquad y_0' = b \qquad [11]$$
$$y'' = 2a \qquad y_0'' = 2a.$$

Therefore, applying [7] and [8], we find

$$y = c + bx + ax^2 \qquad [12]$$

for the individual, and

$$\bar{y} = \bar{c} + \bar{b}x + \bar{a}x^2 \qquad [13]$$

for the group. Thus, the criterion is satisfied and we can consider the averaging a legitimate operation.

If, however, $y = M - Me^{-kx}$, then

$$y_0 = 0$$
$$y' = Mke^{-kx} \qquad y_0' = Mk$$

$$y'' = -Mk^2e^{-kx} \quad y_0'' = -Mk^2 \quad [14]$$
$$y''' = Mk^3e^{-kx} \qquad y_0''' = Mk^3$$
$$\vdots \qquad\qquad \vdots$$

Applying [7] and [8], we find

$$y = 0 + Mkx - Mk^2\frac{x^2}{2!}$$
$$+ Mk^3\frac{x^3}{3!} - \cdots \qquad [15]$$

and

$$\bar{y} = 0 + \overline{(Mk)}x - \overline{(Mk^2)}\frac{x^2}{2!}$$
$$+ \overline{(Mk^3)}\frac{x^3}{3!} - \cdots. \qquad [16]$$

In [16] the values of the coefficients are not simply functions of the average parameters, $\bar{M}$ and $\bar{k}$, but rather averages of the products, Mk, Mk^2, etc., and therefore the averaging process is not legitimate, as has been indicated by Sidman for this function. It will be apparent that the failure of the exponential growth function to satisfy the criterion results from the fact that $\bar{y}$ is dependent not upon the parameters, as such, but upon a variety of *products* of the parameters; and the mean of a series of products is not necessarily the same as the product of the means. Since the average curve depends upon these individual products, no inferences can be made about the nature of the individual functions from the average curve without knowledge of these individual products. Furthermore, if we knew enough about the individual functions to ascertain these individual products, the average curve would be quite gratuitous.

"THE RESPONSE TO COLOR AND EGO FUNCTIONS": A CRITIQUE IN THE LIGHT OF RECENT EXPERIMENTAL EVIDENCE

J. D. KEEHN

Institute of Psychiatry, Maudsley Hospital, London

In a recent article in this Journal Fortier (5) set out a theory intended to establish a correspondence between color and affect. Beginning from and integrating the theoretical views of Schachtel (12) and Rickers-Ovsiankina (11), Fortier went on to present a number of publications allegedly supporting the view that response to color is indicative of emotionality. It is the purpose of this note to examine briefly the validity of the evidence brought forward in support of this theory.

The pertinence of the views of Schachtel and Rickers-Ovsiankina need not concern us here. Suffice it to say that they attempt logically to equate color with affect. Instead, it is more important to examine the experimental evidence purporting to throw light on the efficacy of this logic.

Three main lines of experimental evidence are cited: (a) evidence from the Rorschach test, (b) evidence from the Mosaic Test, and (c) evidence from examination of finger paintings.

In view of the overwhelming majority of studies quoted by Fortier making use of the Rorschach, this paper is confined to an examination of these studies. This is not unreasonable if one remembers that finger paintings and the Mosaic Test are in no way standardized and therefore cannot be regarded as of sufficient validity for the adequate testing of a theory. As Fortier finds no correlation between the results obtained from the Rorschach and those from the Mosaic Test, thus being forced into the construction of a subsidiary ad hoc hypothesis, this restriction should act in his favor.

Examination of the evidence presented shows that all the experiments and observations quoted make use of the principle of external validation. Thus, different groups of varying degrees and modes of "emotionality" were used and the relationship between "emotionality" and Rorschach color score indicated. However, in many cases, instead of finding the expected relationship Fortier finds results not in accordance with his views. Thus, when he finds that Rorschach investigations do not support Burt's (2) view on the emotionality of delinquents, Fortier discards this view for another *which is based upon Rorschach findings*, and so finds an impressive correspondence between the new view of the affective state of the delinquent and his color score on the Rorschach. Similarly, when he finds that Rorschach findings do not fit in with the traditional view, he discards the orthodox view of the emotional sterility of the schizophrenic in favor of Beck's (1) opinion that those patients do have emotional experiences. All this despite the fact that Beck's opinion is largely derived from his experience with the Rorschach test. There are other instances of similar *ad hoc* reasoning, but to continue would be pointless.

65

Even with the genetic studies Fortier finds it difficult to prove his point and complains that little evidence is available. But were Fortier acquainted with the works of Lindberg (9), Eysenck (4), and others he would find numerous papers quoted demonstrating the decrease of color reaction with increasing age in children, evidence which would enable him to accept the general view of the increase of emotional control with age. This evidence comes from experiments with tests of color-form attitude, a line of approach which Fortier seems unaware of, for there is no mention of it in his text, although one reference does appear in the bibliography (Oeser [10]). However, no studies using the external criterion of "emotionality" are of much use in attempting to find a personality correlate of color reactivity, for emotionality itself is but a vague and ill-defined term.

If one wishes to discover whether or not reaction to color in general does have some correlate in the field of personality, then some indirect approach must be used. One such approach would be that of internal validation. This involves administering a number of tests involving color as a scorable item and correlating the resultant test scores. Unless significant correlations were obtained between such tests, color reactivity could not be regarded as having any personality correlates, for if a particular personality attribute were associated with high reaction to color then this reaction would have to occur on all tests involving color reaction as a scorable response.

Examination of the data of Eysenck (4), Clarke (3), and Lindberg (9) (reworked) shows an average correlation between color-form tests of the order of .2. This leads the first two

of these authors to conclude that, except in the eventuality of isolating a factor of color-form attitudes, reactions to color-form tests are relatively specific to each test. Lindberg reached the opposite conclusion through a false statistical analysis, but the reanalysis of his data as shown in Keehn's study (8) supports the opinions of Eysenck and Clarke.

However, a factor analysis of the intercorrelation matrix obtained from a number of color-form tests given to a large group of subjects did result in the extraction of a color-form factor (8). Evidence presented from the same analysis (7) showed that the conventional Rorschach color score had no saturation on this factor but, instead, had a loading on a factor of whole-part attitude.[1] Thus the results from this study were interpreted as an indication that color was used as a determinant in the Rorschach only if the subject attempted to integrate as much of the stimulus as possible into his response. This shows the traditional view of the role of color in the Rorschach test to be incorrect.

Support for the need of some reinterpretation of the part played by color in this test comes from an analysis of the experiments carried out to determine the validity of the concept of color shock. This concept is also relevant to the color-emotionality hypothesis, being, in effect, an extension of it paralleling the constriction of emotion with the suppression of color responses. In a review of a number of studies with color shock (6) hardly any evidence for the dependence of color shock signs upon color was found.

Thus, and this is the crux of the

[1] This evidence, of course, was not available to Fortier during the preparation of his paper.

matter, inasmuch as the Rorschach color responses are at best only indirectly affected by color, the use of the results of Rorschach studies either to support or to refute the color-emotionality hypothesis is invalid. Hence the great majority of the evidence put forward by Fortier in support of his theory must be regarded as irrelevant.

REFERENCES

1. BECK, S. Personality structure in schizophrenics. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.

2. BURT, C. *The young delinquent.* London: Univer. of London Press, 1948.

3. CLARKE, A. D. B. The measurement of emotional instability by means of objective tests. Unpublished doctor's dissertation, Univer. of London, 1950.

4. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.

5. FORTIER, R. H. The response to color and ego functions. *Psychol. Bull.*, 1953, 50, 41–63.

6. KEEHN, J. D. Rorschach validation. II: The validity of colour shock in the diagnosis of neuroticism. *J. ment. Sci.*, 1953, 99, 224–234.

7. KEEHN, J. D. Rorschach validation. III: An examination of the role of colour as a determinant in the Rorschach test. *J. ment. Sci.*, 1953, 99, 410–438.

8. KEEHN, J. D. An experimental investigation into the validity of the use of colour and form attitudes in the assessment of personality. Unpublished doctor's dissertation, Univer. of London, 1953.

9. LINDBERG, B. J. Experimental studies of colour and non-colour attitude in school children and adults. *Acta Psychiat. Neurol.*, 1938, Suppl. 16.

10. OESER, O. Some experiments on the abstraction of form and colour: II. Rorschach tests. *Brit. J. Psychol.*, 1932, 22, 287–323.

11. RICKERS-OVSIANKINA, MARIA. The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, 17, 227–257.

12. SCHACHTEL, E. On color and affect; contributions to an understanding of Rorschach's test. II. *Psychiatry*, 1943, 6, 393–409.

AN APPRAISAL OF KEEHN'S CRITIQUE OF "THE RESPONSE TO COLOR AND EGO FUNCTIONS"

ROBERT H. FORTIER[1]
Northampton, Massachusetts

There are essentially three main criticisms levied by Keehn:

1. *Criticisms based upon Keehn's recently completed work.* It becomes difficult to formulate a reply to criticisms derived from original work which has not yet been published. What can be said here is based only upon the brief excerpt which Keehn included in his critique. Keehn

[1] Now at the VA Hospital, Northampton, Mass.

states: " . . . inasmuch as the Rorschach color responses are at best only indirectly affected by color the use of Rorschach studies either to support or to refute the color-emotionality hypothesis is invalid." The path which he follows from his studies to this conclusion is rather difficult to follow, particularly as it applies to the writer's theoretically oriented review.

Keehn does make two direct state-

ments in his critique before arriving at the rather sweeping conclusion quoted above: (a) "This [referring to his color-form factor analysis] shows the traditional view of the role of color in the Rorschach test to be incorrect." (b) ". . . hardly any evidence for the dependence of color shock signs upon color was found."

In regard to the first statement, the writer is largely in agreement with Keehn that the traditional view of the role of color in the Rorschach test is in many respects incorrect. Such was the point and purport of his entire paper. However, an evaluation of this statement by Keehn could have been better made if Keehn had stated its rationale more specifically. In regard to Keehn's findings of color shock and its relevance to the role of color, it can be seen by a scrutiny of the writer's paper (2) that in no whit does the theory he presents depend upon any interpretation of color shock.

If the two statements quoted here are the vehicles which carried Keehn to his final conclusion, it can be seen that his conclusion has no relevance to the writer's paper.

Keehn, upon receipt of my appraisal of his critique, very kindly sent me a slightly revised copy of the critique, in which he stresses somewhat more the interpretation of his color-form factor analysis. This revision, in turn, prompted me into dealing more directly with these interpretations. They are as follows:

a. Reactions to color-form tests are relatively specific to each test.

b. "Thus, the results from this study were interpreted as an indication that color was used as a determinant in the Rorschach only if the subject attempted to integrate as much of the stimulus as possible into his response."

In regard to *a*, the writer wishes to point to what Keehn has called this writer's "subsidiary *ad hoc* hypothesis" concerning the Mosaic Test. The writer, in that formulation—whatever it may later be called—stressed that the specific response to, or mode of usage of, color is dependent upon the inherent dynamics of the Mosaic Test—or the functions which the individual had to perform in executing the test. Color was still related to affect in the manner in which it was sketched in the theory. The fairly strong implication was that the specific responses elicited by one test might be different from responses elicited by another test simply because the operations satisfying the requirements of the two tests might be different.

In regard to *b*, if affects can be regarded as part of the totality of any situation, that individual who is regarded as more capable of handling affects is the one who can best integrate affects into the totality of the situation, and can control them.

2. Failure to pursue the line of attack developed by Lindberg and others. The original bibliography compiled by the writer consisted of several hundred articles. Limitations of space prevented a meaningful discussion of some of these studies. Other studies were ambiguous or otherwise unclear. In an original draft of his review paper, the writer included the particular study by Lindberg (3) which Keehn cited. However, he found that a great many of the interpretations and conclusions made by Lindberg were completely immersed in a system of psychology which had not appeared to any great extent in the American literature. He felt that a fair treatment of Lindberg's interpretations and conclusions could not be made without dealing at

great length with the particular psychological system used by Lindberg. Because space was limited, such an elaboration was impossible, and Lindberg's study was dropped.

3. *The utilization of* ad hoc *hypotheses.* It was the writer's intent "to present a theory . . . and second, to examine a number of studies involving color in an effort to substantiate and clarify the theory" (2, p. 41). Although this theory has been in existence for a number of years, it is by no means in common usage. Following this presentation, the writer reviewed a number of studies to test the theory. In its strict sense, the term "test" would have implied in the situation under discussion a mere listing of studies with the words "yes" and "no" indicating whether the study supported the theory or refuted it. And following such a listing, one could perhaps have taken a "vote" by counting the "yeses" and "noes." The writer felt that much more could be gained by subjecting the data to a logical analysis, particularly when interpretations and conclusions from the data were contradictory to prevalent points of view.[2] In other words, particularly in those cases where interpretations and conclusions from data were contrary to prevalent views, the writer made

reinterpretations of the original data, stated the reinterpretations, and subsequently attempted to examine their implications. The reinterpretations and inferences derived therefrom thus become "*ad hoc*" in the sense that these formulations were not made until the studies were seen, and their conclusions suggested the need for reinterpretation. Benjamin has this to say concerning *ad hoc* constructs or explanations: "Explanation consists in the transfer of properties and relations from the explanatory entity to that which is to be explained; in explanation one increases the content of the data by adding to them the features of the hypothesis. But if a construct has been obtained by direct derivation from the data, there is clearly no possibility of turning about and explaining the data in terms of the construct. The construct does not contain that increment of novelty without which explanation is impossible. A pure construct used as an explanatory entity produces an *ad hoc* or *verbal* explanation" (1, p. 186). The writer drew his reinterpretations and inferences in many cases from areas which lay outside the specific data reported in the studies which he reviewed. Thus, in the sense in which the term is used by Benjamin, the reinterpretations are not *ad hoc*. The value of Keehn's criticism would have been enhanced had he specified the sense in which he felt these reinterpretations to be *ad hoc*.

---

[2] It is the writer's opinion that his adaptation of Schachtel's and Rickers-Ovsiankina's theories is not among the prevalent points of view, and he so stated in his paper.

REFERENCES

1. BENJAMIN, A. C. *An introduction to the philosophy of science.* New York: Macmillan, 1937.
2. FORTIER, R. H. The response to color and ego functions. *Psychol. Bull.*, 1953, **50**, 41–63.
3. LINDBERG, B. J. Experimental studies of colour and non-colour attitude in school children and adults. *Acta Psychiat. Neurol.*, 1938, Suppl. 16.

CURRENT INTERPRETATION AND SIGNIFICANCE OF LLOYD MORGAN'S CANON

EDWARD NEWBURY
University of Kentucky

In 1932 Nagge (**14**) showed that the meaning of Lloyd Morgan's Canon had been greatly transformed by later psychologists so that, permitting a variety of interpretations, it was difficult to apply and needed clarification. In some instances more recently, as in the earlier paper of Adams (**1**), the Canon has been interpreted as a version of the law of parsimony (Boring, **2**, pp. 474, 498, 623, 629; Griffith, **4**, p. 322; Harriman, **5**, p. 255; Hilgard, **6**, p. 343; Munn, **12**, p. 2; **13**, p. 2; Skinner, **17**, p. 4; Warden, Jenkins, and Warner, **20**, p. 26; Waters, **22**, p. 24; Woodworth, **25**, pp. 49 f.). In several of these (**2**, pp. 563, 623; **5**, **12**, **13**, **25**) and others (**18**, p. 250; **24**, pp. 124, 747), the Canon has been explicitly interpreted as a doctrine of simplicity. It has been related to Occam's Razor (**2**, p. 498; **4**, **5**, **25**), has had imputed to it an antimentalism (**4**, **5**, **17**), or what amounts to a behavioristic reductionism (**17**), but also a futile subjectivism (**20**, pp. 45–54). The Canon has been criticized on several of these bases. This variety of interpretations poses a problem in both accuracy and the historical continuity of thought. What *did* Morgan say?

The Canon and its context. Lloyd Morgan was mainly concerned with the continuity of mental evolution. His Canon was a methodological rule for placing animals in the scale of development. The first statement (1894) was: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (**7**, p. 53). In a revised edition Morgan added to this a clarifying restatement: "In no case is an animal activity to be interpreted in terms of higher psychological processes, if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development" (**8**, p. 59). In 1925 he extended the range of the Canon to embrace his now explicit emergent evolution: "In no instance should we interpret events in terms of concepts appropriate to a higher level of emergence if they can adequately be interpreted in terms of concepts appropriate to a lower level of emergence" (**10**, p. 61). And in his American autobiography (1932) he speaks of "the evolutionary canon that we should not interpret an earlier and lower stage of mental development in terms applicable only to the interpretation of the higher and later stage" (**11**, p. 262; cf. also pp. 260 ff.).

The original Canon was set in a context of double-aspect monism. Introspection was necessary for direct acquaintance with psychical processes. Double inductions were required for inferring mental processes in other organisms, either human or animal, in which "inductions reached through the objective study of certain physical manifestations have to be interpreted in terms of in-

ductions reached through the introspective study of mental processes" (7, p. 47, cf. pp. 36–53).

The context was also one of an evolutionary naturalism which recognized the appearance of novel functions in a process called "selective synthesis" (7, ch. 19). "Synthesis, with new properties at critical turning-points, was the burden of my evolutionary contention," Morgan later commented, changing the term to "emergence" (9, p. 302 ff.).

While Morgan recognized that a given process might be more highly developed in one animal type than in another, he was not willing to accept a reductionism whereby "all forms of animal life from the amoeba upwards have all the faculties of man," or that "in the higher forms of life the introduction of the higher faculties has been effected by some means other than that of natural evolution" (7, p. 58). He argued that "any animal may be at a stage where certain higher faculties have not yet been evolved from their lower precursors; and hence we are logically bound not to assume the existence of these higher faculties until good reasons shall have been shown for such existence" (p. 59). Thus the Canon, in Morgan's conception, was one of the devices displacing an uncritical assumption of identity between the human and animal mind.

The context of the Canon shows that Morgan was not trying to explain the behavior of a single animal in a single experimental situation, but to determine what characteristics of consciousness could be attributed to a given animal type. In his later clarification, "lest the range of the principle be misunderstood," Morgan carefully adds "that the canon by no means excludes the interpretation of a particular activity in terms of the higher processes, if we already have independent evidence of the occurrence of these higher processes in the animal under observation" (8, p. 59). Currently Hilgard makes a similar point, but erroneously gives it as an *inversion* of the Canon. "As now inverted," he writes, "it might be paraphrased: 'An organism capable of ideational problem-solving may also use ideas in learning situations in which they would be theoretically unnecessary" (6, p. 345).

Antimentalism and reductionism. Despite its introspective basis, the Canon has been interpreted as eliminating conscious processes through an antimentalism or by implying in it a reductionism. Harriman quotes Occam's Razor—"entities must not be multiplied beyond necessity"— and states that "Morgan accepted this view, indicating that anecdotes, attribution of human mental activities to animals, and projection of introspections have no place in animal psychology" (5, p. 255). Griffith makes the Canon, interpreted on the basis of Occam's Razor, appear to emphasize physical factors and not to necessitate the inclusion of mental ones (4, p. 322). Skinner writes: "Darwin, insisting upon the continuity of mind, attributed mental faculties to subhuman species. Lloyd Morgan, with his law of parsimony, dispensed with them in a reasonably successful attempt to account for characteristic animal behavior without them" (17, p. 4). And Munn is not clear (13, pp. 1 f.).

Waters with some approval says that "rigid application of Morgan's Canon" is criticized by Gestalt psychologists on the ground of need for "qualitative descriptions that are not allowed on the basis of this canon" (22, pp. 24 f.). Yet the Canon was consistent with the concept of emer-

gent novelties and the configurational qualities such as depth perception in which "the out-thereness" results from sensations coalescing "into a synthesis which has a new and determinate character" (7, p. 352; cf. also ch. 1). The criticism seems to be of a misinterpretation rather than of a rigid application of the Canon.

Waters (21) attacks an anti-anthropomorphism which he ascribed ambiguously either to the Canon or to later interpretations of it. What Waters in general seems to find necessary is a critical anthropomorphism in the sense of introspectively oriented concepts. Since this was the actual basis for Morgan's Canon, the value of historical continuity of thought is lost in Waters' treatment, all the more emphatically because he designates Tolman, Krechevsky, and Maier as anthropomorphic in this sense (pp. 535 f.). Similarly, Morgan's current significance is lost when Hilgard follows a historical treatment by his own criticism of the requirement that explanations be only in palpable terms (6, pp. 343 f.). And the humanistic, textbook aims of Stagner and Karwoski are hardly realized by stating that Morgan "attacked 'anthropomorphism' (interpreting animal behavior in human terms) with his famous canon" (18, p. 250). The Canon is a device for selecting in interpretation introspectively derived processes, not for completely eliminating them, which would be to eliminate use of the Canon (cf. 11, p. 261).

Occam's Razor and parsimony. The antimentalistic contradiction of applying Occam's Razor through the Canon was noted above in Griffith and Harriman. In the sense of Harriman's literal translation (*supra*) of the classical Latin statement (19),

Occam's Razor is applied when we adhere to a *paucity* of assumptions, whereas Morgan's Canon refers to *lower* processes of development. The total number of scientific entities is not decreased by the Canon in so far as they are introspectively available, but the number of emergent functions attributed to an animal of a given type may be decreased, so that for this animal the net effect is in line with Occam's Razor, while the admonition in the Canon regards the lowness of the functions, not their paucity.

As an application of the law of parsimony and of Occam's Razor, Boring argues that however useful the Canon may have been in counteracting extremes of Romanes' anthropomorphism, times and problems have changed, and entities of necessity must now be increased (2, pp. 474, 498, 563). Since the Canon allows higher processes when independent evidence makes their assumption necessary, one may reasonably ask, why not further decrease probability of error by utilizing knowledge of developmental sequences?

While Occam's Razor and the law of parsimony have no doubt sometimes been identified in science, the latter especially is a loose term subject to a variety of meanings, some of which go as far back as Aristotle (2, pp. 498 f.; 3, p. 231; 14; 15, p. 472; 16, pp. 114 f.; 19; 23, p. 155). It is only confusing to say, without adequate qualification, that Morgan's Canon is derived from either, and of doubtful accuracy to identify it with them. The confusion can be compounded by the doctrine of simplicity.

Simplicity of explanation. Woodworth criticizes Morgan's Canon as a doctrine of simplicity, raising the question of what is simpler. Animal

psychologists have usually regarded movement as simpler than perception, he says, but "judging by ourselves we should say that nothing is easier than seeing an object and that no learned reaction is simpler than recognition" (24, p. 124).

Such a criticism was anticipated by Morgan himself (7, p. 54). He answered, first, that "surely the simplicity of an explanation is no necessary criterion of its truth," and in many cases "the simplest explanation is not the one accepted by science" (pp. 54 f.). Second, "the simplicity of the explanation of the phenomena of animal activity as the result of intellectual processes, can only be adopted on the assumption of a correlative complexity in the mental nature of the animal as agent" (p. 55). This assumption had to be justified by induction. Thus Morgan distinguished between the doctrine of simplicity and his Canon, which makes no reference to it.

Morgan did suppose that in evolution the earlier physical and mental processes were simpler (p. 8). He postulated an evolutionary development with an increasing complexity of organic structure, of correlated activities, and of mental and psychic functions (p. 55). This made the problem of psychic level which logically justified the Canon. But in this

development appeared new properties. Indeed, years later Morgan turned his 1925 version of the Canon on the radical behaviorist by saying that "we should not interpret the lower forms of trial and error as affording evidence of trial and error of the kind that characterises reflective procedure" (10, p. 61). Thus a doctrine of higher and lower in the level of development can be consistent with the presumptive fact of evolutionary development from simplicity to complexity without the methodological rule being one of simplicity at all. Failure to recognize this possibility seems to explain a good deal of the misinterpretation of the Canon.

Comment. Aside from their historical inaccuracies, many current misinterpretations of Morgan's Canon have *sui generis* failed to take advantage of possible logical developments. Without contending that Morgan's methodology represents the last word, one can recognize in it some of the essentials for integrating modern introspective and comparative psychology. Whether this gain through historical continuity can be realized depends upon an accurate and significant interpretation of that methodology, including the Canon.

REFERENCES

1. ADAMS, D. K. The inference of mind. *Psychol. Rev.*, 1928, 35, 235–252.
2. BORING, E. G. *A history of experimental psychology.* (2nd Ed.) New York: Appleton-Century-Crofts, 1950.
3. CURTIS, S. J. *A short history of Western philosophy in the Middle Ages.* London: Macdonald, 1950.
4. GRIFFITH, C. R. *Principles of systematic psychology.* Urbana: Univer. of Illinois Press, 1943.
5. HARRIMAN, P. L. *The new dictionary of psychology.* New York: Philosophical Library, 1947.
6. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
7. MORGAN, C. L. *An introduction to comparative psychology.* London: Walter Scott, 1894.
8. MORGAN, C. L. *An introduction to comparative psychology.* (2nd Ed. Rev.) London: Walter Scott, 1906.
9. MORGAN, C. L. *Emergent evolution.* New York: Holt, 1923.
10. MORGAN, C. L. *Life, mind, and spirit.* New York: Holt, 1925.
11. MORGAN, C. L. Autobiography. In C. Murchison (Ed.), *A history of psychol-*

ogy in autobiography. Vol. II. Worcester, Mass.: Clark Univer. Press, 1932. Pp. 237–264.

12. MUNN, N. L. *Psychological development.* Boston: Houghton Mifflin, 1938.

13. MUNN, N. L. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950.

14. NAGGE, J. W. Regarding the law of parsimony. *J. genet. Psychol.,* 1932, 41, 492–494.

15. PEARSON, K. *The grammar of science.* London: Walter Scott, 1892.

16. RITCHIE, A. D. *Scientific method.* New York: Harcourt, Brace, 1923.

17. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

18. STAGNER, R., & KARWOSKI, T. F. *Psychology.* New York: McGraw-Hill, 1952.

19. THORBURN, W. M. Occam's razor. *Mind,* 1915, N.S., 24, 287–288.

20. WARDEN, C. J., JENKINS, T. N., & WARNER, L. H. *Comparative psychology.* Vol. 1. New York: Ronald, 1935.

21. WATERS, R. H. Morgan's canon and anthropomorphism. *Psychol. Rev.,* 1939, 46, 534–540.

22. WATERS, R. H. Historical background of comparative psychology. In C. P. Stone (Ed.), *Comparative psychology.* (3rd. Ed. Rev.) New York: Prentice-Hall, 1951.

23. WEYL, H. *Philosophy of mathematics and natural science.* Princeton: Princeton Univer. Press, 1949.

24. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

25. WOODWORTH, R. S. *Contemporary schools of psychology.* (Rev. Ed.) New York: Ronald, 1948.

PSYCHOLOGICAL NECROLOGY (1928-1952)

SUZANNE BENNETT AND EDWIN G. BORING

Harvard University

It is not easy to find the dates of death of prominent men because the who's-who compendia usually omit the sketch when the man dies and thus never publish the date. For this reason one of us published in 1928 the dates of death of 248 prominent psychologists who died during the twenty-five year period 1903–1927 (1). Now, a quarter of a century later, we undertake for the same reason to present a similar list for the next twenty-five years, 1928–1952. How did we get the names and data?

We searched for this period *Psychological Abstracts*, the *Directories of the American Psychological Association*, *American Men of Science*, *Science*, and *Archives de Psychologie*, which was especially rich in European notices during Claparède's incumbency. We added, of course (and continued later to add), various prominent names which came incidentally to our attention, and we came out with a list of 809 deceased psychologists, one death every twelve days for a quarter century! Since not all these persons were important enough to list in crowded journal space, we had to select.

So we gave the list of 809 to L. M. Terman, H. S. Langfeld, Wayne Dennis, J. G. Beebe-Center, E. B. Newman, and the senior author of this paper, asking them to indicate which names they did not recognize at all, which names they recognized without further association, which names they recognized and knew something about, as well as the names they recognized but thought too unimportant for inclusion in the final list and the names they recognized but thought deviated too far in the sociotropic, biotropic, or physicotropic direction to count as psychologists. Then we edited these results, making decisions.

We omitted such prominent names as Ruth Benedict, Bergson, Franz Boas, C. B. Davenport, Hrdlicka, Korzybski, Malinowski, Raymond Pearl, and J. J. Thomson, persons who are or were important to psychologists but not, in our opinion, quite psychologists. We allowed ourselves, as our list shows, to deviate more freely in the direction of physiology and neurology, in accordance with psychology's tradition in which *physiologische Psychologie* was more important than *Völkerpsychologie*. So we include Cajal, H. H. Donaldson, Ferrier, Flechsig, W. M. Wheeler, and some other biotropes. We let in sociotropes as remote as Havelock Ellis, and the important psychoanalysts, like Adler, Ferenczi, and Rank. Freud, of course. Nor could we quite bring ourselves to exclude Karl Pearson.

Next we had regard to the votes of our jury of six. We accepted for the list every name, not already excluded as peripheral to psychology, that received six, five, or four votes as being known about in addition to being barely recognized. That gave us nearly all of our list. Finally we accepted 12 names with only three votes, mostly names where once famous psychologists were known to

the older members of the jury and not to the younger. The senior author then translated his prejudices into authority and took in three persons with only two votes and Beatrice Edgell whose contribution to psychology was recognized by none of the jury other than himself. Nobody was included on bare recognition alone. Thus we obtained a list of 193 persons, to which we have added gratuitously entries for three important psychologists who died early in 1953: Kafka, Katz, and Marbe.

The list is published for the purpose of recording exact dates of death. Instead of giving age at time of death we have included in all cases exact date of birth. It has not been easy to obtain these data. We have consulted over two dozen files, from the *New York Times* and *Who Was Who in America* to the published necrologies in the standard scientific journals. For the last score of entries we wrote letters to persons who might know about the deceased or have ready access to local records. Professor P. Fraisse of Paris and Professor H. v. Bracken of Braunschweig have been particularly helpful in obtaining some of the French and German items.

Our geographical entry is meant for identification. It is not the place of death, which is sometimes an obscure locality to which an old man had retired. We have tried to give what seemed to us to be the place of last important professional activity.

Here is the list.

Ach, Narziss Kasper, Göttingen, b. 7 June 1848, d. 25 July 1946.

Adler, Alfred, Vienna and New York City, b. 7 Feb. 1870, d. 28 May 1937.

Adler, Herman Morris, Univer. California, b. 10 Oct. 1876, d. 6 Dec. 1935.

Angell, Frank, Stanford Univer., b. 8 July 1857, d. 2 Nov. 1939.

Angell, James Rowland, Yale Univer., b. 8 May 1869, d. 4 Mar. 1949.

Angier, Roswell Parker, Yale Univer., b. 21 Oct. 1874, d. 24 June 1946.

Arps, George Frederick, Ohio State Univer., b. 23 Jan. 1874, d. 16 Sept. 1939.

Aveling, Francis, London, b. 25 Dec. 1875, d. 6 Mar. 1941.

Babcock, Harriet, New York City, b. 7 Jan. 1877, d. 17 Dec. 1952.

Bagley, William Chandler, Columbia Univer., b. 15 Mar. 1874, d. 1 July 1946.

Baldwin, Bird Thomas, Univer. Iowa, b. 31 May 1875, d. 12 May 1928.

Baldwin, James Mark, Paris, b. 12 Jan. 1861, d. 8 Nov. 1934.

Becher, Erich, Munich, b. 1 Sept. 1882, d. 5 Jan. 1929.

Bingham, Walter Van Dyke, Washington, D. C., b. 29 Oct. 1880, d. 8 July 1952.

Bleuler, Paul Eugen, Zürich, b. 30 Apr. 1857, d. 15 July 1939.

Bolton, Thaddeus Lincoln, Temple Univer., b. 27 July 1865, d. 3 Jan. 1948.

Book, William Frederick, Indiana Univer., b. 10 June 1873, d. 22 May 1940.

Bourdon, Benjamin Bienaimé, Rennes, b. 5 Aug. 1860, d. 11 July 1943.

Breese, Burtis Burr, Univer. Cincinnati, b. 17 May 1867, d. 31 July 1939.

Brett, George Sidney, Univer. Toronto, b. 5 Aug. 1879, d. 27 Oct. 1944.

Brigham, Carl Campbell, Princeton Univer., b. 4 May 1890, d. 24 Jan. 1943.

Brill, Abraham Arden, Columbia Univer., b. 12 Oct. 1874, d. 2 Mar. 1948.

Buckner, Edward Franklin, Johns Hopkins Univer., b. 3 Sept. 1868, d. 22 Aug. 1929.

Burks, Barbara Stoddard, Columbia Univer., b. 22 Dec. 1902, d. 25 May 1943.

Burnham, William Henry, Clark Univer., b. 3 Dec. 1855, d. 25 June 1941.

Burrow, Trigant, Lifwynn Found., b. 7 Sept. 1875, d. 24 May 1950.

Cajal, Santiago Ramón y, Madrid, b. 1 May 1852, d. 17 Oct. 1934.

Calkins, Mary Whiton, Wellesley Coll., b. 30 Mar. 1863, d. 26 Feb. 1930.

Cannon, Walter Bradford, Harvard Univer., b. 19 Oct. 1871, d. 1 Oct. 1945.

Cason, Hulsey, Univer. Miami, b. 21 Feb. 1893, d. 30 Apr. 1951.

Cattell, James McKeen, New York City, b. 25 May 1860, d. 20 Jan. 1944.

Claparède, Edouard, Geneva, b. 24 Mar. 1873, d. 29 Sept. 1940.

Coghill, George Ellett, Wistar Inst., b. 17 Mar. 1872, d. 23 July 1941.

Conklin, Edmund Smith, Indiana Univer., b. 19 Apr. 1884, d. 6 Oct. 1942.

Coover, John Edgar, Stanford Univer., b. 16 Mar. 1872, d. 19 Feb. 1938.

Dearborn, George Van Ness, New York City, b. 15 Aug. 1869, d. 12 Dec. 1938.

Decroly, Ovide Jean, Brussels, b. 23 July 1871, d. 12 Sept. 1932.

Delabarre, Edmund Burke, Brown Univer., b. 25 Sept. 1863, d. 16 Mar. 1945.

Delacroix, Henri, Paris, b. 2 Dec. 1873, d. 3 Dec. 1937.

De Sanctis, Sante, Rome, b. 7 Feb. 1862, d. 20 Feb. 1935.

Dessoir, Max, Berlin, b. 8 Feb. 1867, d. 19 July 1947.

Dewey, John, Columbia Univer., b. 20 Oct. 1859, d. 1 June 1952.

Dockeray, Floyd Carlton, Ohio State Univer., b. 15 May 1880, d. 15 Jan. 1949.

Dodge, Raymond, Yale Univer., b. 20 Feb. 1871, d. 8 Apr. 1942.

Donaldson, Henry Herbert, Wistar Inst., b. 12 May 1857, d. 23 Jan. 1938.

Downey, June Etta, Univer. Wyoming, b. 13 July 1875, d. 11 Oct. 1932.

Drever, James, Edinburgh, b. 8 Apr. 1873, d. 11 Aug. 1950.

Dumas, Georges, Paris, b. 6 Mar. 1866, d. 13 Feb. 1946.

Duncker, Karl, Swarthmore Coll., b. 2 Feb. 1903, d. 23 Feb. 1940

Dunlap, Knight, Univer. California Los Angeles, b. 21 Nov. 1875, d. 14 Aug. 1949.

Dusser de Barenne, Johannes Gregorius, Yale Univer., b. 6 June 1885, d. 9 June 1940.

Edgell, Beatrice, London, b. 26 Oct. 1871, d. 10 Aug. 1948.

Ehrenfels, Christian von, Prague, b. 20 June 1859, d. 8 Sept. 1932.

Ellis, Henry Havelock, London, b. 2 Feb. 1859, d. 8 July 1939.

Farrand, Livingston, Cornell Univer., b. 14 June 1867, d. 8 Nov. 1939.

Ferenczi, Sándor, Budapest, b. 7 July 1873, d. 22 May 1933.

Fernald, Grace Maxwell, Univer. California Los Angeles, b. 29 Nov. 1879, d. 16 Jan. 1950.

Fernald, Mabel Ruth, Cincinnati, Ohio, b. 7 May 1883, d. 9 Oct. 1952.

Ferree, Clarence Errol, Johns Hopkins Univer., b. 11 Mar. 1877, d. 26 July 1942.

Ferrier, David, London, b. 13 Jan. 1843, d. 19 Mar. 1928.

Flechsig, Paul Emil, Leipzig, b. 29 June 1847, d. 22 July 1929.

Fletcher, John Madison, Tulane Uni-

ver., b. 27 June 1873, d. 12 Dec. 1944.

Forel, August, Zürich, b. 1 Sept. 1848, d. 27 July 1931.

Franz, Shepherd Ivory, Univer. California Los Angeles, b. 27 May 1874, d. 14 Oct. 1933.

Freud, Sigmund, Vienna and London, b. 6 May 1856, d. 23 Sept. 1939.

Frey, Max von, Würzburg, b. 16 Nov. 1852, d. 25 Jan. 1932.

Fröbes, Joseph, Rome, b. 26 Aug. 1866, d. 24 Mar. 1947.

Gamble, Eleanor Acheson McCulloch, Wellesley Coll., b. 2 Mar. 1868, d. 30 Aug. 1933.

Garth, Thomas Russell, Univer. Denver, b. 24 Dec. 1872, d. 20 Apr. 1939.

Geissler, Ludwig Reinhold, Randolph-Macon Women's Coll., b. 22 Sept. 1879, d. 15 Dec. 1932.

Gelb, Adhémar, Frankfurt a. M., b. 18 Nov. 1887, d. 7 Aug. 1936.

Gilliland, Adam Raymond, Northwestern Univer., b. 5 Oct. 1887, d. 1 Dec. 1952.

Goldscheider, Alfred, Berlin, b. 4 Aug. 1858, d. 10 Apr. 1935.

Gosset, William Sealy ("Student"), London, b. 13 June 1876, d. 16 Oct. 1937.

Groos, Karl, Tübingen, b. 10 Dec. 1861, d. 3 Apr. 1946.

Grünbaum, Anton Abram, Utrecht, b. 23 May 1885, d. 10 Jan. 1932.

Haggerty, Melvin Everett, Univer. Minnesota, b. 17 Jan. 1875, d. 6 Oct. 1937.

Haines, Thomas Harvey, Montclair, N. J., b. 4 Nov. 1871, d. 2 Mar. 1951.

Hamilton, Gilbert Van Tassel, Santa Barbara, Calif., b. 15 Jan. 1877, d. 16 Dec. 1943.

Head, Henry, London, b. 4 Aug. 1861, d. 8 Oct. 1940.

Hecht, Selig, Columbia Univer., b. 8 Feb. 1892, d. 18 Sept. 1947.

Henmon, Vivian Allen Charles, Univer. Wisconsin, b. 27 Nov. 1877, d. 10 Jan. 1950.

Heymans, Gerardus, Groningen, b. 17 Apr. 1857, d. 18 Feb. 1930.

Hobhouse, Leonard Trelawney, London, b. 8 Sept. 1864, d. 21 June 1929.

Höffding, Harald, Copenhagen, b. 11 Mar. 1843, d. 2 July 1931.

Hollingworth, Leta Stetter, Columbia Univer., b. 25 May 1886, d. 27 Nov. 1939.

Holt, Edwin Bissell, Princeton Univer., b. 21 Aug. 1873, d. 25 Jan. 1946.

Holway, Alfred Harold, Eastman Kodak Co., b. 28 Feb. 1905, d. 1 June 1948.

Hornbostel, Erich von, Berlin and Cambridge, Eng., b. 25 Feb. 1877, d. 13 June 1935.

Horney, Karen, New York City, b. 16 Sept. 1885, d. 4 Dec. 1952.

Hull, Clark Leonard, Yale Univer., b. 24 May 1884, d. 10 May 1952.

Husserl, Edmund, Freiburg i. Br., b. 8 Apr. 1859, d. 27 Apr. 1938.

Isaacs, Susan Sutherland, London, b. 24 May 1885, d. 12 Oct. 1948.

Jaensch, Erich Rudolf, Marburg, b. 26 Feb. 1883, d. 1 Feb. 1940.

Janet, Pierre, Paris, b. 30 May 1859, d. 24 Feb. 1947.

Jastrow, Joseph, Univer. Wisconsin, b. 30 Jan. 1863, d. 8 Jan. 1944.

Jelliffe, Smith Ely, New York City, b. 27 Oct. 1866, d. 25 Sept. 1945.

Jenkins, John Gamewell, Univer. Maryland, b. 30 May 1901, d. 30 Jan. 1948.

Jennings, Herbert Spencer, Johns Hopkins Univer., b. 8 Apr. 1868, d. 14 Apr. 1947.

Judd, Charles Hubbard, Univer. Chicago, b. 20 Feb. 1873, d. 18 July 1946.

Kafka, Gustav, Würzburg, b. 23 July 1883, d. 12 Feb. 1953.

Katz, David, Stockholm, b. 1 Oct. 1884, d. 2 Feb. 1953.

Kellogg, Chester Elijah, McGill Univer., b. 11 Nov. 1888, d. 9 July 1948.

Kiesow, Federico, Turin, b. 28 Mar. 1858, d. 9 Dec. 1940.

Kirkpatrick, Edwin Asbury, Fitchburg St. Teach. Coll., b. 29 Sept. 1862, d. 4 Jan. 1937.

Kirschmann, August, Leipzig, b. 21 July 1860, d. 24 Oct. 1932.

Klemm, Gustav Otto, Leipzig, b. 8 Mar. 1884, d. 5 Jan. 1939.

Koffka, Kurt, Smith Coll., b. 18 Mar. 1886, d. 22 Nov. 1941.

Kries, Johannes von, Freiburg i. Br., b. 6 Oct. 1853, d. 30 Dec. 1928.

Kuhlmann, Frederick, St. Paul, Minn., b. 20 Mar. 1876, d. 19 Apr. 1941.

Ladd-Franklin, Christine, Columbia Univer., b. 1 Dec. 1847, d. 5 Mar. 1930.

Le Bon, Gustave, Paris, b. 7 May 1841, d. 14 Dec. 1931.

Leuba, James Henry, Bryn Mawr Coll., b. 9 Apr. 1868, d. 8 Dec. 1946.

Lévy-Bruhl, Lucien, Paris, b. 10 Apr. 1857, d. 12 Mar. 1939.

Lewin, Kurt, Mass. Inst. Tech., b. 9 Sept. 1890, d. 12 Feb. 1947.

Lindley, Ernest Hiram, Univer. Kansas, b. 2 Oct. 1869, d. 21 Aug. 1940.

Lindworsky, Johannes, Prague, b. 21 Jan. 1875, d. 9 Sept. 1939.

Link, Henry Charles, Psych. Corp., b. 27 Aug. 1889, d. 9 Jan. 1952.

Lipmann, Otto, Berlin, b. 6 Mar. 1880, d. 7 Oct. 1933.

Lipps, Gottlob Friedrich, Zürich, b. 6 Aug. 1865, d. 9 Mar. 1931.

MacCurdy, John Thomson, Cambridge, Eng., b. 18 Jan. 1886, d. 1 July 1947.

MacDougall, Robert, New York Univer., b. 12 June 1866, d. 1 Nov. 1939.

McDougall, William, Duke Univer., b. 22 June 1871, d. 28 Nov. 1938.

McGeoch, John Alexander, Univer. Iowa, b. 9 Oct. 1897, d. 3 Mar. 1942.

Marbe, Karl, Würzburg, b. 31 Aug. 1869, d. 2 Jan. 1953.

Marston, William Moulton, New York City, b. 9 May 1893, d. 2 May 1947.

Martin, Lillian Jane, San Francisco, Calif., b. 7 July 1851, d. 26 Mar. 1943.

Maxfield, Francis Norton, Ohio State Univer., b. 29 Aug. 1877, d. 10 Nov. 1945.

Mead, George Herbert, Univer. Chicago, b. 27 Feb. 1863, d. 26 Apr. 1931.

Meyer, Adolf, Johns Hopkins Univer., b. 13 Sept. 1866, d. 17 Mar. 1950.

Miner, James Burt, Univer. Kentucky, b. 6 Oct. 1873, d. 24 Mar. 1943.

Monakow, Constantin von, Zürich, b. 4 Nov. 1853, d. 19 Oct. 1930.

Morgan, Conwy Lloyd, Bristol, b. 6 Feb. 1852, d. 6 Mar. 1936.

Morgan, John Jacob Brooke, Northwestern Univer., b. 23 Aug. 1888, d. 16 Aug. 1945.

Müller, Georg Elias, Göttingen, b. 20 July 1850, d. 23 Dec. 1934.

Myers, Charles Samuel, London, b. 13 Mar. 1873, d. 12 Oct. 1946.

Patrick, George Thomas White, Univer. Iowa, b. 19 Aug. 1857, d. 21 May 1949.

Pavlov, Ivan Petrovitch, Leningrad, b. 26 Sept. 1849, d. 27 Feb. 1936.

Pearson, Karl, London, b. 27 Mar. 1857, d. 24 Feb. 1936.

Peterson, Joseph, Peabody Coll., b. 8 Sept. 1878, d. 20 Sept. 1935.

Pintner, Rudolf, Columbia Univer., b. 16 Nov. 1884, d. 7 Nov. 1942.

Poppelreuter, Walther, Bonn, b. 6 Oct. 1886, d. 11 June 1939.

Prince, Morton, Boston, Mass., b. 21 Dec. 1854, d. 31 Aug. 1929.

Pyle, William Henry, Wayne Univer., b. 27 Feb. 1875, d. 3 Mar. 1946.

Rank, Otto, New York City, b. 22 Apr. 1884, d. 31 Oct. 1939.

Ranschburg, Paul, Budapest, b. 3 Jan. 1870, d. 18 Jan. 1945.

Richet, Charles, Paris, b. 26 Aug. 1850, d. 3 Dec. 1935.

Rignano, Eugenio, Milan, b. 31 May 1870, d. 9 Feb. 1930.

Robinson, Edward Stevens, Yale Univer., b. 18 Apr. 1893, d. 27 Feb. 1937.

Robinson, Florence Richardson, New Haven, Conn., b. 3 July 1885, d. 3 Dec. 1936.

Rosanoff, Aaron Joshua, Sacramento, Calif., b. 26 June 1878, d. 7 Jan. 1943.

Rubin, Edgar John, Copenhagen, b. 6 Sept. 1886, d. 3 May 1951.

Ruch, Giles Murrel, Washington, D. C., b. 7 July 1892, d. 15 Nov. 1943.

Sachs, Hanns, Boston, Mass., b. 10 Jan. 1881, d. 10 Jan. 1947.

Saudek, Robert, London, b. 21 Apr. 1881, d. 15 Apr. 1935.

Schiller, Paul Harkai, Yerkes Lab. Primate Biology, b. 4 Nov. 1908, d. 1 May 1949.

Schumann, Friedrich, Frankfurt a.M., b. 16 June 1863, d. 10 Mar. 1940.

Seashore, Carl Emil, Univer. Iowa, b. 28 Jan. 1866, d. 16 Oct. 1949.

Seashore, Robert Holmes, Northwestern Univer., b. 14 June 1902, d. 27 Aug. 1951.

Shand, Alexander Faulkner, London, b. 20 May 1858, d. 7 Jan. 1936.

Sharpey-Schafer, Edward Albert, Edinburgh, b. 2 June 1850, d. 29 Mar. 1935.

Shepherd, William Thomas, Columbia Univer. Sch., Washington, D. C., b. 4 Feb. 1867, d. 2 Feb. 1945.

Sherrington, Charles Scott, Oxford b. 27 Nov. 1857, d. 4 Mar. 1952.

Shirley, Mary Margaret, Indiana Univer., b. 7 Feb. 1899, d. 11 June 1946.

Small, Willard Stanton, Univer. Maryland, b. 24 Aug. 1870, d. 31 Jan. 1943.

Smith, Stevenson, Univer. Washington, b. 29 Apr. 1883, d. 27 Nov. 1950.

Snoddy, George Samuel, Indiana Univer., b. 21 Apr. 1882, d. 29 June 1947.

Spearman, Charles Edward, London, b. 10 Sept. 1863, d. 17 Sept. 1945.

Starbuck, Edwin Diller, Univer. Southern California, b. 20 Feb. 1866, d. 18 Nov. 1947.

Stern, William, Duke Univer., b. 29 Apr. 1871, d. 27 Mar. 1938.

Stetson, Raymond Herbert, Oberlin Coll., b. 1 Mar. 1872, d. 4 Dec. 1950.

Stoelting, Christian H., Chicago, Ill., b. 18 July 1864, d. 18 Mar. 1943.

Stout, George Frederick, St. Andrews, b. 6 Jan. 1860, d. 18 Aug. 1944.

Stumpf, Carl, Berlin, b. 21 Apr. 1848, d. 25 Dec. 1936.

Sutherland, Arthur Howard, Louisiana Coll., b. 19 Nov. 1878, d. 18 May 1951.

Swift, Edgar James, Washington Univer., b. 24 July 1860, d. 30 Aug. 1932.

Tait, William Dunlop, McGill Univer., b. 20 Nov. 1879, d. 10 May 1945.

Thorndike, Edward Lee, Columbia Univer., b. 31 Aug. 1874, d. 9 Aug. 1949.

Triplett, Norman, Kansas St. Teach. Coll., Emporia, b. 1 Oct. 1861, d. 16 Oct. 1934.

Troland, Leonard Thompson, Harvard Univer., b. 26 Apr. 1889, d. 27 May 1932.

Twitmyer, Edwin Burket, Univer. Pennsylvania, b. 14 Sept. 1873, d. 3 Mar. 1943.

Valentine, Willard Lee, Northwestern Univer., b. 2 Dec. 1904, d. 5 Apr. 1947.

Warren, Howard Crosby, Princeton Univer., b. 12 June 1867, d. 4 Jan. 1934.

Washburn, Margaret Floy, Vassar Coll., b. 25 July 1871, d. 29 Oct. 1939.

Weiss, Albert Paul, Ohio State Univer., b. 15 Sept. 1879, d. 3 Apr. 1931.

Wembridge, Eleanor Harris Rowland, Los Angeles, Calif., b. 9 Dec. 1883, d. 20 Feb. 1944.

Wertheimer, Max, New School for Social Research, b. 15 Apr. 1880, d. 12 Oct. 1943.

Wheeler, William Morton, Harvard Univer., b. 19 Mar. 1865, d. 19 Apr. 1937.

Whipple, Guy Montrose, Clifton, Mass., b. 12 June 1876, d. 1 Aug. 1941.

White, William Alanson, St. Elizabeth's Hosp., Washington, D. C., b. 24 Jan. 1870, d. 7 Mar. 1937.

Willoughby, Raymond Royce, Providence, R. I., b. 20 Apr. 1896, d. 3 Oct. 1944.

Wirth, Wilhelm, Leipzig, b. 26 July 1876, d. 13 July 1952.

Woolley, Helen Bradford Thompson, Columbia Univer., b. 6 Nov. 1874, d. 24 Dec. 1947.

Yoakum, Clarence Stone, Univer. Michigan, b. 11 Jan. 1879, d. 20 Nov. 1945.

Yule, George Udny, Cambridge, Eng., b. 18 Feb. 1871, d. 26 June 1951.

Zwaardemaker, Hendrik, Utrecht, b. 10 May 1857, d. 19 Sept. 1930.

REFERENCE

1. BORING, E. G. Psychological necrology (1903–1927). *Psychol. Bull.*, 1928, 25, 302–305, 621–625.

SPECIAL REVIEW[1]

TEXTBOOKS AND GENERAL PSYCHOLOGY

FRANK W. FINGER

University of Virginia

ASHER, ESTON J., TIFFIN, JOSEPH, & KNIGHT, FREDERIC B. *Introduction to general psychology.* Boston: Heath, 1953. Pp. xvi+515. $4.25.

BUXTON, CLAUDE E., COFER, CHARLES N., GUSTAD, JOHN W., MACLEOD, ROBERT B., McKEACHIE, WILBERT J., & WOLFLE, DAEL. *Improving undergraduate instruction in psychology.* New York: Macmillan, 1952. Pp. vii+60. $1.25.

COLE, LAWRENCE E. *Human behavior: Psychology as a bio-social science.* Yonkers-on-Hudson: World Book Co., 1953. Pp. xi+884. $5.50.

GUILFORD, J. P. *General psychology.* (2nd Ed.) New York: Van Nostrand, 1952. Pp. xii+587. $5.00.

HILGARD, ERNEST R. *Introduction to psychology.* New York: Harcourt, Brace, 1953. Pp. x+659. $5.75.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953. Pp. x+461. $4.00.

STAGNER, ROSS, & KARWOSKI, T. F. *Psychology.* New York: McGraw-Hill, 1952. Pp. xiii+582. $5.00.

The temerity of the textbook writer who believes that he has produced the answer to general psychology's prayer is exceeded only by the audacity of the reviewer who presumes to pass judgment on the belief. More rash than wise is the individual who would go still further and attempt an evaluative comparison of several current texts. It may be fun to predict consumer reaction, or to dare some hardy instructor to edge himself out on an untested textual limb. But a more helpful function of the review is to present the new books in broad context, leaving the responsibility for evaluation where it belongs—with each individual teacher. In view of the theoretical differences that characterize psychology, and the diversities in philosophy and practice typifying the teaching profession, it would be folly to expect that any teaching aid could be all things to all pedagogues.

If there is any discernible trend running through all higher education today, it is a healthy compulsion to self-examination. It is interesting to speculate that its temporal proximity to World War II and continuing international crisis is more than coincidental. Manpower problems are more obviously critical, the success of selection and training methods in the armed services suggests that parallel problems in the academic field might yield to equivalent experimental analysis, and perhaps there is impatience with a traditional scholarship that has failed to open the road to Utopia. Whatever the explanation may be, the consequence is clear. Spurred on by Presidential commissions and educational foundations, every college institutes some new curricular reform, or at least renames an old one; private and public soul-searching is the order of the day.

In view of psychology's subject

matter and demonstrated skills, it would be anomalous indeed if we were unwilling to confront the situation in our own bailiwick. While much of our extra-laboratory energy has been applied to clinical and industrial problems, curriculum rationale and teaching procedures have been receiving increasing amounts of attention. It is becoming gradually accepted that graduate departments may with profit organize seminars to guide their students in the development of instructional proficiency. The Division on the Teaching of Psychology is growing not only in membership but in productive activity. Most important, the recognition is spreading that criteria are fully as elusive in the classroom as in pilot training, that sampling biases are not automatically eliminated by college matriculation, and that pedagogical preconceptions, like experimental hypotheses, exist to be challenged.

What's Wrong with General Psychology?

The focus of discussion has been the introductory psychology course. It might be argued that the disproportionate importance of this course makes it worthy of greatest effort, for it serves the greatest number of students and it is the primary medium of recruitment. Further, it probably encompasses the problems of all but the most technical courses, both undergraduate and graduate. And, if further excuse is needed, the beginning is the logical place to begin. What shall be our objectives in this course: to increase the student's appreciation of the advantages and limitations of scientific method, to nourish the development of more objective attitudes toward human behavior, to enhance general understanding of behavior dynamics by delineating basic psychological principles, to give a balanced view of what constitutes the scientific discipline and the professional practice called psychology, to help the student solve his emotional and vocational problems? Should we lecture to our classes, or plan directed discussions, or just sit and exude warm acceptance? Is the major assigned reading to repeat the class activities or is it to cover the topics for which no class time is available? Is the number of pages apportioned to the several topics to reflect the relative interest that a sample of unsophisticated students have expressed? Is the reading to furnish case material for group dissection? Is it to add the living flesh of research to the bare skeleton of topics orally presented? Is it to be entertaining and lively, or sober and pedestrian?

A Guide: Buxton et al.

A partial set of answers comes from the symposium on the improvement of undergraduate psychology instruction, held at Cornell during the summer of 1951. Much of the report deals with the introductory course. It is first assumed that the desires and expectations of the students should not dictate the nature of the course, although they may well be exploited to facilitate the learning process. Problems of everyday interest may furnish the point of departure, but the instructor must be sure to depart a perceptible distance. This is definitely not to be a personal adjustment course, and the intellectual goals must never become secondary to the urge to conduct group therapy. The primary teaching emphasis in the ideal curriculum is to be upon knowledge and content (the problems of psychology, facts

and principles of psychology, psychology as science, the structure and functioning of science), which should then indirectly lead the student toward the objectives of rigorous habits of thought and modified attitudes toward behavior.

The study group finds that the effectiveness of the usual introductory text is impaired by the unfortunate combination of two divergent organizations, and proposes that either the developmental or the cross-sectional approach be explicitly adopted and rather strictly followed. In the sample outlines presented, less than the traditional emphasis is placed upon the topics of neural anatomy and physiology, the receptors, the effectors, tests, individual differences, and statistics. While no selection and arrangement of topics can remain definitive in a changing field, the adoption of some guiding organization is preferable to a listless eclecticism born of inertia and ancestor worship.

Science and Practical Living: Hilgard

Hilgard's basic orientation involves two major objectives, which at first glance may seem to be incompatible. He wishes the text to give a fair view of psychology as a whole, and to introduce the student to "those topics which are the centers of excitement among professional psychologists today" (p. ix). At the same time, he recognizes as valid the desires of the elementary student to receive help in increasing self-understanding. To reconcile these aims is not impossible, he feels. After all, more psychologists are actively concerned with industrial and clinical matters than with retinal interaction. And so it is fair enough to devote nearly 20 per cent of the book to a section entitled "Psychology Applied to Personal and Social Problems" ("Mental Health and Readjustment Techniques," "Vocational Adjustment," "Psychology in Industry," "Public Opinion and Propaganda," "Problems of Social Groups"), while omitting the usual chapters on receptor structure and processes, the nervous system, biological genetics, and statistics. Still more significant is the impression that Hilgard is striving to be "person-oriented" in the other 80 per cent, most obviously so in the chapters on "Adolescence and Adulthood," "Emotion and Motivation," "Social Motives," "Conflict and Frustration," and "Individual Modes of Adjustment." He is not simply describing a series of representative experiments from the relevant sections of *Psychological Abstracts*. He is instead analyzing some major problems of behavior in terms understandable to the student, and recounting what he has come to think about them after all these years as a psychologist; experiments are interpolated chiefly to indicate that his thinking is not entirely autistic.

If relegation of the laboratory to the relatively minor position will please the student, it will just as surely disappoint some instructors. This is first of all a course in *science*, chiefly *experimental* science. Is it possible to preserve the spirit of research in a 16-half-line abbreviation of a 32-page experimental article? Can six or seven pages toward the end of the book, formally describing scientific procedure, adequately acquaint the student with the fundamental problems and limitations of experimental methodology and inculcate in him the desired habits of rigorous thinking? Will not the essay form, the soft pedaling of controversial data, give a false impression of ease and certainty in scien-

this discovery? Is it enough merely to state the final polished results of investigations? What of the endless struggle with experimental controls and the slow refinement of quantitative measurement? Or can we assume that a competent instructor will be an indispensable half of the course? (It is almost paradoxical that of the four really new texts reviewed, this is the one that least needs a constant interpreter and fellow discussant, with its delightfully readable and yet mature style.)

A Unifying Concept: Stagner and Karwoski

It is often suggested that the best preparation for the Ph.D. comprehensive examination is to study an introductory text. Stagner and Karwoski's *Psychology* is admirably suited for this use. They make no attempt to be easy, to skip over difficult concepts, to subordinate the standard subject matter of psychology to matters of immediate practicality, or to "write down." While this may satisfy the graduate student's needs, it may by the same token discourage the sophomore. If the services of the professor are required to add rigor to Hilgard they will be helpful here to assure understanding and everyday interest. The one-semester course based on this text can be very solid but will surely not be easy.

A frequently voiced complaint is that general psychology is a loose stringing together of disparate topics. It is difficult for the instructor to depict for the student an integrated organism unless some unifying concept is invoked. It is to homeostasis that Stagner and Karwoski turn to fulfill this function. Some degree of success is achieved by this device. But the tying together of topics is

accomplished still more meaningfully in another fashion. There is laudable willingness to mention a concept outside the chapter or the chapter section where it "belongs." Thus, the longish chapter labeled "Perception" involves considerations of personality, motivation, social factors; and perception recurs in the discussion of problem solving, of remembering, of thinking, of intelligence, of personality. Such contextual cross-referencing of course fosters continuing expansion of the meaning of each idea. Other writers, as well as teachers, would do well to extend this demonstration of the basic integrity of the behaving organism.

Revisions: Guilford; and Asher, Tiffin, and Knight

These two books, based as they are on earlier publications, represent the least drastic shift from the established pattern of the past twenty or thirty years. They will perhaps fit most readily into the traditional course organization.

In 1941 Guilford was reported a best seller. Unless your favorite experiment is among those pruned away, there is little reason for you to like this blue printing any less than the red printing. A dozen pages have been added on "What Psychologists Do," there is a new 30-page section on heredity and development, and the rest of the material is presented in somewhat modified order.

Asher, Tiffin, and Knight, in what is essentially its third edition, has shifted its center of gravity somewhat toward the right—slightly farther removed from the strictly applied and closer to psychology as science. Of the 14 chapters in this relatively short book three are new ("Development of the Individual," "Body Structures and Behavior,"

and "Motivation"), and others have been partially rewritten.

A Cross-Discipline Approach: Cole

It is not always easy to judge whether an author's second book in a field is different enough from the first to require additional shelf space. A simple criterion now suggests itself: if another publisher is involved, any similarity between the two is not only coincidental but probably illegal! *Human Behavior* represents a considerable departure from Cole's earlier *General Psychology* (2) and from all other elementary psychology texts, in both emphasis and organization.

The most startling difference is found in the concluding section, 200 pages devoted to the "self-system." After a chapter relating the hypnotic state and the divided self, there are five chapters describing major aspects of psychoanalytic theory. The final chapter, "The Normal Personality," is a 50-page discourse which, as Cole suggests, goes beyond psychology in answering the "So what?" of the entire course. The reader will look in vain for the usual discussion of types vs. traits, an outline of personality tests, and the classification of deviations. Instead he will find an original dissertation on the philosophical and practical meaning of "normality," with an unusually varied selection of allusions to literature, theoretical psychiatry, and theology. It will be a strong dose for the student to swallow as he comes to the end of his survey of psychology. The average instructor will have to be shaken loose from a great many scientific fixations if he is to regard the treatment as nourishing rather than purgative.

The second distinguishing feature is the abundance of documentation throughout the book, drawing freely and at length from experimental descriptions, philosophical treatises, sociological and anthropological expositions, and a wide variety of case reports. If Hilgard tends to use illustrations as footnotes to his discussion, and if Stagner and Karwoski incorporate their references more intimately into the textual material, Cole can be said in some chapters to be providing the interpretive continuity for an extended anthology of behavior stories. The average beginning student will find that a great deal of the material makes interesting reading, but he will probably also find himself in need of considerable classroom guidance if he is to abstract the pattern of principles usually considered to constitute the core of general psychology. In a very real sense, we have another difficult text.

Surprise Entry: Skinner

Inclusion of this item in the review may be unexpected and even controversial. Admittedly it would be impossible if the list were restricted to books deliberately designed for the general psychology course or to those deviating no more than one sigma from the current norm. But even if it is unacceptable for widespread classroom adoption, its consideration should force us to scrutinize a little more closely whatever faith underlies our practice in the first course.

Science and Human Behavior was prepared in connection with the Harvard general education course in behavior science. The unhappy condition of the world is attributed to a failure to resolve, or even to face, a basic contradiction between opposed conceptions of man. "A scientific conception of human behavior dic-

tates one practice, a philosophy of personal freedom another" (p. 9). A wise choice presupposes understanding of the alternatives; it is Skinner's responsibility to clarify the first position, by providing a meaningful definition of science, by demonstrating that human behavior is amenable to the manipulations of experimental science, and by considering the implications for practical living that must inevitably emerge. His belief that the variables significant for the understanding and control of behavior lie only in the immediate environment and the behavioral history of the organism is illustrated in 150 pages of relatively nontechnical and nonquantitative discussion of operant behavior. Where the laboratory demonstration involves pigeons—and it not infrequently does—logically parallel examples from common human experience are proposed. For example, the cumulative effect of accidental contingencies in operant feeding reminds us of human superstition, the unusual efficacy of certain reinforcement schedules seems to be relevant to the optimal payoff rate of gambling devices, verbal responses are modified under the influence of discriminative stimuli, imitative birds have their counterpart in human dancers, punishment temporarily suppresses the rate but leaves unchanged the rat's or child's total number of responses. Throughout this section, the systematic position of the writer is never obscure. (Tip to the graduate student: these chapters should be reprinted as "Skinner Almost Painlessly Revealed.")

To reinforce the premise that complex behavior is within the potential range of science, the principles of simple cases are then extended to self-control, to thought processes and inner events, and to the functionally unified system of responses sometimes called "the self." The social situation is brought within the same framework, beginning with the two-person relationship and relentlessly (perforce speculatively) driving on to the controlling agencies of government, religion, psychotherapy, economics, and education. Like Cole, Skinner recognizes that when the science of behavior is pushed to its ultimate, the problem of who is to control whom and for what ends runs squarely up against some of the most cherished conceptions of human life. He does not counsel retreat.

How does this help us in introductory psychology? How can we develop in our students a respect for psychology as a quantitative and controlled mode of investigation if we omit the details of experimental design and procedure, if we fail to talk in terms of means and deviations? Can anything but a distorted view result from trying to survey modern experimental psychology from the confines of the Skinner box? What happens to all the time-honored observations and principles that we have come to know as general psychology?

Perhaps the answer is that these facts and generalizations are unimportant and soon forgotten, that clear perspective can most economically be gained from a consistent vantage point, and that numbers and gadgets are not the essence of science. Surely the story of the science-behavior marriage in the typical text is by contrast to this exposition pale and unsubstantial, the attempt to relate psychology to other aspects of life pitifully tentative. For those who feel that this part of our job is the most important, a little more of the Skinnerian approach would seem to

be in order. If the unsupplemented classroom use of *Science and Human Behavior* is precluded by limitations of curriculum and clientele, its assimilation by the serious teacher is most appropriate.

Trends

Does the curve pass through enough points to permit extrapolation?

To summarize the contemporary trends in the general psychology course would require a very complex formula indeed, with about as many negative terms as positive. And yet some hints may emerge from a comparison of the present set of texts with a somewhat earlier reference group, such as Woodworth and Marquis (6), Dashiell (4), Cruze (3), and Munn (5).

Topics. The most obvious difference is in the line-up of chapter and section headings. There seems to be a tendency, though not unopposed, to transfer the formal study of statistical techniques and the organic bases of behavior to advanced courses or to other departments. Thus statistics covers a total of perhaps six pages in the six books; the nervous system as a separate topic appears only in the two revisions; genes have all but disappeared from the new books; Hilgard and Cole find it possible to write general texts without chapters on the receptors. With the addition of development sections to the revisions, only Stagner and Karwoski deviate in this respect from the reference texts. And yet any listing of labels or counting of pages must be viewed with suspicion; for example, it is doubtful that any of the books with the development chapter qualify as pure representatives of the "developmental" orientation outlined by Bux-

ton *et al.*, and it would be incorrect to assume that any but Skinner would exclude neurophysiological variables from psychology. It is more nearly correct to say that a trend away from compartmentalization has blurred certain arbitrary distinctions and encouraged the juxtaposition of items formerly separated in writing. So, Stagner and Karwoski interject a discussion of significant characteristics of the nervous system at any point where it may clarify other topics, and Cole calls his approach *bio*social in spite of his chapter omissions.

While convincing quantitative proof may be lacking, I have the impression that more emphasis is now being placed upon personality, and especially upon its social determinants. The self concept has come into prominence (Hilgard, Stagner and Karwoski, Cole, Skinner) and there is no question but that the insights of psychoanalysis are enjoying a much more general acceptance. This, together with Hilgard's and Guilford's chapters surveying the fields of psychological endeavor, may suggest that general psychology, like the APA, is becoming concerned with the profession as well as the science.

The whimsical order of topics has traditionally been a source of variety to lighten the yearly chore of teaching introductory psychology. There is in the present sample good evidence that some measure of stability is being approached—at least, development and motivation come early, and personality late.

Specific content and format. More deadly to the chronic instructor than a fixed list of chapters is the same diet of illustrative stories, experiments, and pictures regurgitated from one generation of writers to the next. The student may be unaware

of the difference, but the jaded teacher will be refreshed by some of the new dishes cooked up by the current authors. Hilgard has been the most resourceful in the search for new ingredients to satisfy old recipes, and Stagner and Karwoski's chapter on thinking has a very liberal sprinkling of original elements.

Generosity in the use of pictorial material, as well as variation in size and arrangement of page, is at once an economic, an aesthetic, and a psychological problem. Some publishers must recognize that the principles of satiation and of diminishing returns are as pertinent as the principles of attention. It may be for this reason that the use of half-tones is quite limited in Cole and in Asher *et al.*, and moderate in Stagner and Karwoski, and Guilford. Hilgard, on the other hand, has what might be called a superabundance, and distraction begins to outweigh clarification. It is probably helpful to portray in action representatives of three extreme somatotypes, and one well chosen picture may be more valuable than three hundred words analyzing depth perception, but it is questionable that a picture of ten giraffes or a ship under construction will contribute a half-page worth to the student's understanding of gregariousness or demoralization. The very obvious trend of increasing prices might be decelerated without educational loss if the criteria of moderation and relevance were more prominent in the determination of form.

Passing mention should be made of the excellent bibliography furnished in Hilgard. It contains upwards of 850 items, given in complete form and keyed to the pages in the text on which reference is made. There is also a glossary of about 500 terms.

Systematic bias. There have been very few elementary texts with the controversial flavor of Titchener and Watson, and our first five suggest no unique trend in this regard. Few teachers are blessed with systematic conscience so strict as to prohibit assignment of any of the chapters (except perhaps Cole's critical exposition of Freudianism), although as usual there will be some choking over definitions, distress at internal inconsistencies, and puzzlement as to how experience is to be examined if the instrument of introspection is discarded. Skinner is of course not so universally inoffensive. With all its restrictive disadvantages, this at least can be said of such a persistent theoretical treatment—that the patient knows he has been treated. When the student finishes Skinner (or vice versa), he will be aware that he has been up against something, whether good or bad. Too few college experiences can similarly be characterized.

Teaching aids. It would seem logical that the instructor is in the best position to decide how a textbook and other supplementary devices can optimally be fitted into the course plan. However, as a guide for the inexperienced or preoccupied and as a source of new ideas, many writers furnish extratextual material. Guilford, and Asher *et al.*, include self-test questions after each chapter. Separate manuals for instructor and student accompany both Hilgard, and Stagner and Karwoski. Their quality indicates that greater than usual effort was expended in their preparation. Incidentally, it is in the Stagner and Karwoski material that we find the only suggestion that laboratory work can be a useful part of the introduction to psychology.

The Moral

No great perspicacity is required to sense the restlessness that pervades instruction in general psychology. We can't quite agree on what we should be doing, or how. The problem certainly isn't solved when in grammatical confusion we echo the old wheeze about teaching students instead of psychology. The report of the Cornell symposium does not supply the final word (cf. 1). Nor should we expect any textbook writer to solve the riddle for us. It is not enough, for example, to adopt a text whose preface solemnly asserts that it is "student oriented" and which thereafter departs from the old mold only by addressing the reader as "you" instead of "one."

The answers must be found in the course itself, not in the auxiliary aids provided. There are a dozen texts that are still adequately up to date and that can be used with profit by the capable class and the thoughtful instructor. No one of them can assume the responsibility for teaching (To add to the cliché, you can't "teach a book," either.) Indeed, if a choice were forced as to which locus of student-orientation is the more critical, it should obviously be the day-by-day activity of the instructor, with the printed supplement simply a sound and representative description of psychological science. There are many ways of teaching effectively; it is up to each instructor to generate some hypothesis as to what procedures and what adjuncts are most fruitful, and then test and revise and retest the hypothesis *ad infinitum*. The diversity of textbook offerings is most fortunate, if it has the effect of forcing the instructor into a questioning and experimental frame of mind. The crucial self-examination in higher education is after all at the level of the individual.

REFERENCES

1. BERRIEN, F. K., & CASTORE, G. F. A note on elementary psychology. *Amer. Psychologist*, 1953, 8, 246.
2. COLE, L. E. *General psychology.* New York: McGraw-Hill, 1939.
3. CRUZE, W. W. *General psychology for college students.* New York: Prentice-Hall, 1951.
4. DASHIELL, J. F. *Fundamentals of general psychology.* (3rd Ed.) New York: Houghton Mifflin, 1949.
5. MUNN, N. L. *Psychology. The fundamentals of human adjustment.* (2nd Ed.) New York: Houghton Mifflin, 1951.
6. WOODWORTH, R. S., & MARQUIS, D. G. *Psychology.* (5th Ed.) New York: Holt, 1947.

HULL, CLARK L. *A behavior system: an introduction to behavior theory concerning the individual organism.* New Haven: Yale Univer. Press, 1952. Pp. ix + 372. $6.00.

In the preface to this book, written shortly before his death, Professor Hull places it as the second in a three-volume plan. The first volume, *Principles of Behavior*, appeared in 1943. He writes that it was designed "to state the more important primary principles considered necessary to mediate the deductions of a natural-science theory of behavior." The intent of the present work, as the second volume, is "to show the application of the principles to the deduction of the simpler phenomena characterizing the behavior of single organisms." The plan called for a third volume which would "apply these same principles to the deduction of the elementary phenomena of social behavior, i.e., of behavior manifested when the interacting objects are mammalian organisms of the same species." Professor Hull foresaw that he would probably not live to write the third volume. In some sense this second volume is his final bequest to the science of psychology. As such, it deserves sober review as the crowning achievement of more than 20 years devoted to this type of systematic formulation.

In fairness to Professor Hull, it must be remembered that this work was written during his years of failing health, when he could work but a few hours a day, and when he could not rely, as formerly, on an active and continuing intellectual interchange with his colleagues and students. Hence the critic may detect incompletenesses and contradictions which

perhaps ought not to be there, yet which would not have been there had Professor Hull been able to complete this with the same painstaking care and meticulous revision which characterized his earlier published papers and books. This book should be reviewed in the spirit of one seeking to find out what can be learned from it by way of both method and content; it would be improper to concentrate upon errors of detail. Professor Hull did not believe that he had said the last word. His hope, expressed in the preface and in the conclusion, is that the method exposes its own errors, so that serious students can carry forward the work of making succeeding systems more precise.

Those who know the history of theoretical psychology will understand that the present system is merely the most recent of a series of miniature systems evolved by the present writer. The coming generation of scientists will, it is hoped, present other theoretical systems, each succeeding one of a progressively more precise and quantitative nature (p. 353).

I shall attempt to deal, first, with some of the more ingenious substantive contributions found in the book, second, with its systematic and methodological significance, and, third, with a general estimate of Professor Hull's place in the history of twentieth-century psychology.

Substantive Contributions

Those who were troubled by the limitations of topical coverage in *Principles of Behavior* will be pleased to find in the present volume many of the topics from the papers which appeared in the *Psychological Review* in the early 1930's, including behavior in relation to objects in space, multidirectional maze learning, the

problem-solving assembly of behavior segments.

In planning his deductive system, Hull made many early forays into more complex behavior, including the social behavior that he intended to treat in the final projected volume. But then he retraced his steps, so that, beginning with revised postulates, he could move from simple to more complex behavior. In *Principles of Behavior* he was concerned chiefly with the system of postulates, and so chose to deal quantitatively with behavior near to that formulated in the postulate system, with a minimum of what he called "multiple-link" deductions. The revised postulates were published in a little volume, *Essentials of Behavior* (1951), which chiefly brought the postulates in *Principles of Behavior* up to date as a background for the second volume. Now in *A Behavior System* he reviews, revises, and extends many of the deductions of more complex behavior which he had toyed with before the postulate system was well formulated. Hence, this new volume recovers some of the richness of the early papers which brought his work to the attention of psychologists interested in many aspects of learning remote from classical and instrumental conditioning.

As a brief introduction to some of the substantive achievements, I would direct the reader to the following six deductions, with the corresponding empirical verification:

a. Response alternation in trial-and-error learning. The reference experiment is one in which a rat can be reinforced by moving either a horizontal or a vertical bar. When both bars have been reinforced, and then a shift in reinforcement is made from the more strongly and recently reinforced bar to the other, there ensues a period of alternation before the response is given more frequently to the (now) regularly reinforced bar. The systematic relationships predicted from the theory (Figure 11, p. 48) are borne out empirically with some qualifications (Figure 15, p. 52).

b. Generalization gradients to stronger and weaker stimulus intensities. The theoretical differences in curvature of the gradients (Figure 28, p. 82) is borne out empirically (Figure 29, p. 83).

c. Latent learning. The shifts in reaction with shifts in incentive found in the early studies by Tolman and his collaborators, demonstrating latent learning, are now predicted. The prediction is shown in Figure 37, p. 144, the empirical material in Figure 38, p. 145, and Figure 39, p. 146.

d. Reaction latency at points in a behavior chain. It is predicted that for a four-unit chain, reaction latencies will decrease for three units and increase for the final unit. The prediction is shown in Figure 42, p. 163, the empirical verification in Figure 43, p. 164.

e. Effect of motivation upon behavior in a conflict between an adient and an abient object. The discussion and derivations are found on pp. 245–252, culminating in Theorems 88 and 89, both of which are verified in the empirical outcomes of Table 32, p. 253.

f. Entrances into goalward-pointing alleys vs. those pointing away from the goal, in a multidirectional maze. The prediction of greater frequency of entrances into goalward-pointing blinds is summarized in Theorem 112, p. 289, the empirical data in Table 35, p. 291.

Professor Hull has made his own computation as to agreement between prediction and empirical find-

...ing. His figures are summarized in Table 1.

Curiously enough, the result would be somewhat more convincing were the score poorer, rather than better. The success is not so convincing because too many of the derivations are very close to the data predicted, representing almost a working backward from what is known.

TABLE 1

EMPIRICAL AGREEMENT WITH THEORETICAL PROPOSITIONS (From Hull, *A behavior system*, pp. 351–353)

Propositions	Number	Per Cent
All theoretical propositions		
Evidence found bearing on validity	93	52
Indirect evidence only	30	17
Subtotal	123	69
Not covered by known relevant evidence	55	31
Total	178	100
Propositions covered by direct or indirect evidence		
Substantially validated	106	86
Probably valid; considerable uncertainty	14	11
Definitely invalid	1	1
Not classified	2	2
Total	123	100

Working backwards is not, in itself, a defect in a theoretical system, provided the system is coherent and internally consistent, and provided it also is fertile in predicting relationships not yet explored. As Hull points out, the 55 predictions not covered by known relevant evidence will provide a good test, for these predictions are free of the charge of working backward from the known.

The range of coverage of the system is sufficient to test its systematic and deductive character, for the predictions cover many kinds of behavior not referred to directly in the postulates.

Methodological and Theoretical Contributions

Hull makes ingenious use of the hypothetico-deductive system, in which quantitative treatment moves back and forth between basic principles, or postulates, and empirical consequences, expressed in corollaries; no one can seriously doubt the usefulness of such a model as that which he has furnished. The method is perhaps best elaborated as an exercise in logic and mathematics in the earlier multiple-author volume entitled *Mathematico-Deductive Theory of Rote Learning* (1940). The similar approach, but with a different postulate system, in *Principles of Behavior* and now in *A Behavior System* is a model which will endure in the history of psychology.

The basic condition of learning, according to Hull, is repeated reinforcement which comes about through drive reduction. Some very important changes took place in Hull's quantitative theory of reinforcement between *Principles of Behavior* (1943) and *A Behavior System* (1952). While a complete discussion of these changes would require many pages, some of the main changes may easily be noted.

Habit strength (sH_R) in the earlier book increased with reinforcement. The *amount* of increase per reinforcement depended upon (a) the magnitude of need reduction per reinforcement, (b) the delay in reinforcement, and (c) the interval between the conditioned stimulus and the response to be conditioned to it. In the new book habit strength is solely a function of the number of repetitions of a closely associated stimulus and response, provided this S-R conjunction is reinforced. The amount of reinforcement, and the time relation-

ships, no longer affect $_sH_R$. Hence the basic learning condition becomes, in fact, much closer to that proposed in the familiar views of Guthrie and, more strikingly, of Tolman, because the basic learning process depends upon contiguous S-R relationships, relatively independent of motivation. The difference is preserved in that some minimum of reinforcement must occur. (This necessary minimum is nowhere made clear in the postulates or the text.) The main advantage of this change in relation of $_sH_R$ to need reduction lies in the possibility of deducing latent learning—the empirical basis for Tolman's views in the first place.

What habit strength ($_sH_R$) loses, reaction potential ($_sE_R$) gains. The distinction between habit strength and reaction potential is essentially the distinction between learning and performance, early emphasized by Lashley and Tolman. Habit strength is energized into actual behavior only by being converted to reaction potential through the effect of drive (D), stimulus-intensity dynamism (V), and incentive motivation (K). The delay in reinforcement (J) also affects the size of reaction potential which can lead to response at some link in the behavior chain. The drive (D) was familiar in the 1943 version, but (V) is new and (K) plays a new role. The stimulus intensity dynamism (V) refers to the strength of the signalling (conditioned) stimulus. A stimulus of greater intensity will evoke a greater response, even though $_sH_R$ remains constant. Incentive motivation (K) refers to the quantity of incentive used in earlier trials. It affects $_sE_R$, but not $_sH_R$. (There is no reference in the postulates to the size or nature of the *perceived* incentive in adient or abient situations, an important variable in

much motivated learning.)

With increasing emphasis upon secondary reinforcement, the precise differences between primary and secondary reinforcement become important. In 1943 it is said that primary reinforcement is the result of diminution of a need or drive, although elsewhere it is said that it is the drive stimulus S_D (not D) that is reduced. This latter statement is the one adopted in 1952, with the addition of the possibility that reduction in S_G (the stimulation consequence of the fractional anticipatory goal response) is reinforcing.

If stimulus termination (rather than need reduction) is the basis for reinforcement, then secondary reinforcement is more readily accounted for; it becomes, in 1952, a *corollary* of primary reinforcement rather than a separate postulate. (There are in fact two corollaries of primary reinforcement, one describing secondary motivation as the substitution of an associated neutral stimulus as the condition for a drive, the other describing how a previously neutral stimulus becomes, through association, capable of acting as a reinforcing agent.) I find a certain glibness in the announcement of these corollaries. That something of the sort occurs in learning is clear enough, but precisely what occurs is the focus of controversy between those who favor cognitive theories over reinforcement theories. Any blurring at this point is therefore unfortunate, for it leads to claims of successes in prediction through a kind of *tour de force* by which any conditions which lead to learning must necessarily involve a drive stimulus, and the reduction of the drive stimulus by an associated reinforcing agent. All that needs to be shown is a casual association of some feature of the environment

with aroused drive in the past, and of some feature with reduced drive in the past. The actual linking of stimulus and response intermediaries is often not demonstrated, only their possibilities. Such possibilities exist, of course, in any sign-significate relationship, so that sign learning appears to be easily deduced from these principles. Because only some minimal amount of reinforcement (primary or secondary) is necessary in developing habit strength, the case for a reinforcement theory is made even more plausible. This reinforcement theory requires scarcely any reinforcement, and even that little may be secondary Such a theory is not likely to be successfully refuted, even though it may be false.

It is to Hull's credit that he has searched about for formulations to take into account such demonstrated facts as came to his attention. He was willing to make quite radical modifications in his sytem, as here illustrated by the shift in properties between $_sH_R$ and $_sE_R$. He struggled with the nature of primary and secondary reinforcement. As Spence has pointed out, these modifications which make a great deal of difference in relation to certain controversies among learning theorists actually make rather little difference in the usefulness of a system of this kind, for rather simple modification in the equations commonly takes care of quite different theoretical interpretations.

There is a good deal more of straight empiricism in Hull's method than meets the eye. By comparison with some of the newer mathematical and probabilistic models it is not very mathematical at all, relying heavily as it does upon curve-fitting. This is in some ways a virtue, for these empirical relationships, if derived from careful experiment, will survive a good deal of tampering with basic assumptions about reinforcement.

Hull's Place in Twentieth-Century Psychology

Hull worked out in detail much that was implied in Watson's behaviorism. Taking Pavlov's experiments as a source of his initial postulates, he started, or gave an impetus to, a new kind of system-building in psychology. The older psychological systems, represented in the "schools" of psychology, were largely programmatic, representing ways of defining the field, classifying phenomena, selecting units, and stating certain gross "laws" or principles by which the field could be ordered. While Hull's systematic bias was that of one of these schools (behaviorism), his method was simply that of scientific logic, in which there is an interplay between empirical data and rational ways of dealing with such data. The systematic formulations lead to precise quantitative predictions, so that the system is, to some extent at least, self-correcting. This is the novelty that Hull contributed: a system at once fertile in its predictions, and precise enough to be vulnerable to experimental attack. Because he used a good deal of curve-fitting along the way, the theory has a closer affiliation with functionalism than some of the newer probabilistic models which are more rational (less *ad hoc*) than his.

There is an elegance in the effort that Hull made to develop a limited set of postulates to lead, by strict derivation, to theorems and corollaries subject to experimental test. He held up Newton's mechanics as the ideal, where a few laws of motion could account for the orbits of the planets, the tides, and the path of a projectile. He did not succeed, as

Newton did, but perhaps it is instructive to remember that Newton, too, has been superseded. Hull has set a pattern richly influencing psychology today. He has succeeded in making a permanent place for himself in the history of psychology.

ERNEST R. HILGARD.
Stanford University.

VERNON, M. D. *A further study of visual perception.* New York: Cambridge Univer. Press, 1952. Pp. xi + 278. $7.00.

This book, although similar to the author's *Visual Perception* published in 1937, is not a revised edition but a new book. It is the nearest approximation in print to a comprehensive review of experiments on perception as traditionally distinguished from sensation. That is to say, it omits the evidence from sensory physiology, optics, and ophthalmology, and concentrates on experiments concerning the perception of form and pattern, movement, the perceptual constancies, individual differences in perception, and the facts of attention, attitude, and motivation. There is little overlap, for instance, with Bartley's *Vision*, and it covers some material omitted from Boring's *Sensation and Perception in the History of Experimental Psychology*. It might be said to be a book about *nonstimulus determinants* of perception. The experimental literature as so defined is voluminous, scattered, and contradictory. The facts as such are hard to systematize since they accumulated in the course of theoretical controversies. The author has surveyed a considerable part of this evidence by managing to cover over 500 studies in 260 pages. She has included not only the German but the French experimenters. The most valuable feature of the book may be that it will

cause investigators to look up a great many references they might otherwise have neglected.

The work of Michotte and his students on the frontier between psychophysics and cognition is little understood in this country and Vernon's chapter on these experiments should prove especially useful. No comparable account exists in English. There are also good reviews of the perception of movement, of the time error in psychophysical judgment, of color constancy, and of the problems allied to attention—a word which Vernon is not afraid to use despite the controversies over its meaning.

Vernon's debt to Professor Bartlett of Cambridge is acknowledged at the outset, and her theoretical approach is allied to his. She appears to make approximately the following assumptions: first, that the perceived field is radically unlike the stimulus field— the percept is "immensely modified" as compared with the sensory stimulation; second, that sensory impulses are the raw material for perception— percepts have to be constructed; third, that the individual constructs his perceptual world in accordance with his past experience and his personality—his attitude toward an object is an intrinsic part of his perception of it. The nature of this construction process is, then, the crucial question for theory. A good many writers in addition to Vernon are struggling with this question today, especially those who hope to find in the process a key to the personality of the perceiver. So long as one assumes that perception is based on sensation it is an inescapable question. Vernon shares with other contemporaries a certain disillusion with the explanation by Gestalt organization, which seems less convincing to her than it did in 1937.

Nevertheless she, like others, uses the terminology of organization theory and has not explicitly rejected it. In the end she concludes that we do not yet understand the inner directing tendencies or schemata which shape our percepts.

The failure of the evidence to cohere, however, may be the fault of the basic assumptions under which it was gathered. 'Perhaps the construction process in perception is hard to discover because it does not exist. Perhaps the objective determination of perception and the subjective determination of perception cannot be mixed in a single theory of normal perception; they are incommensurable and what is required is a separate theory for each. Perhaps we must distinguish between *literal* and *schematic* perception. The reviewer has proposed a special theory of the first type but Vernon disagrees, believing that perception is always schematic, and has added an appendix defending her emphasis on subjective determination.

JAMES J. GIBSON.
Cornell University.

SHNEIDMAN, EDWIN S., JOEL, WALTHER, & LITTLE, KENNETH B. *Thematic test analysis.* New York: Grune & Stratton, 1951. Pp. xi+ 320. $8.75.

Shneidman's primary purpose in *Thematic Test Analysis* was to compile a comprehensive manual of the various scoring schemes now in use for analyzing TAT stories and to demonstrate the application of such schemes to a given clinical case. A secondary purpose was to compare the relative merits of the TAT and the MAPS test for the prediction of actual behavior. To these ends he enlisted the aid of fifteen TAT methodologists, requesting them to score and interpret the TAT and MAPS stories of one John Doe. A separate chapter is devoted to each of the analyses. Behavioral data, psychiatric observations, and other test data—Wechsler-Bellevue, Rorschach, Bender-Gestalt, Minnesota Multiphasic, and Draw-A-Person— are included for the reader but were not generally available to the individual analysts. The book ends with a brief synthesis and summary by Shneidman and his collaborators, Joel and Little. The final summary chapter presents a résumé of the different approaches and their classification into five categories: normative (Cox and Sargent, Eron, Hartmann, Klebanoff), intuitive (Bellak, Holt, Lasaga, Rotter and Jessor, Symonds), interpersonal (Arnold, Joel and Shapiro, White), and perceptual (used by none as a primary technique, but by several as a secondary technique). Tabulation and quantitative analysis are left for a second book. Shneidman concludes with two "rather striking impressions" (p. 307): (1) The various clinicians show "remarkable" agreement with each other (p. 307), and "they correlate quite well with the behavioral data" (p. 303). (2) There is a commendable spirit of objective inquiry and personal humility prevailing among clinicians; they are willing "to take a position, right or wrong and to run the risk of public scrutiny" (p. 307). Neither of these conclusions will satisfy the scientifically oriented reader. To correlate "quite well" is statistically meaningless, especially in the absence of quantification. "To run the risk of public scrutiny" is nothing new to the scientist. The book in its present form serves primarily as an illustration of how various clinicians would score the same set of thematic productions.

Much of the real work of synthesis and most of the evaluation is left to the reader.

Granted that a great deal can be learned from a careful study of many different approaches to scoring thematic productions, how are we to evaluate the contributions made by this particular compilation of materials? For the student who wishes to become proficient in a given scoring system, not enough details are given by most of the analysts. Some score all of the stories, others only one or two. Moreover, many of the schemes depend so heavily on the intuition and background knowledge of the analyst that very little can be gained from the sketchy working notes which have been included. What we have here would be better termed a brief survey of scoring schemes, not a manual.

A more basic problem for evaluation centers on Shneidman's own methodology. Here the toughminded experimentalist will find much to trouble him. He could fairly object to the lack of scientific rigor in the research plan and to the plan itself. Just what, he could ask, is being examined here? Is the ultimate object of the research to test the relative usefulness of two different sets of projective materials, TAT and MAPS, for the blind diagnosis of one John Doe; or are the relative merits of the several different scoring schemes per se the major concern? Neither purpose is adequately served by the data assembled for analysis.

The data, stories told by a 25-year-old unmarried male patient to eleven TAT pictures and seven MAPS settings, were not obtained under comparable conditions. The MAPS test was administered before insulin shock therapy was undertaken, the TAT two months later by a different examiner. Behavioral data on John Doe were made available to one scorer, at her request, not to the others. What the effects of these differences are is an interesting problem of experimental study. Yet the analysts were apparently unaware that such differences existed.

Another serious shortcoming is that individual scoring schemes were devised for the TAT, but not for the MAPS test. Many of the analysts recognize that their individual scheme is not entirely suited to the MAPS test. What profit is there, then, in forcing the stories into an ill-fitting mold? Again, most of the analysts are plagued by the absence of adequate norms for the TAT. For the MAPS they have no norms at all. Some indicate that they are attempting to establish norms for the TAT. Others are content merely to apply their own clinical knowledge to the interpretation of John Doe's stories. For communicating a scoring scheme to others, of course, this has serious limitations.

The problem of quantifying the results of the analyses is further complicated by the nature of the individual analyses since the schemes vary from fairly precise scores based on checklists (e.g., Fine) and rating scales (e.g., Eron, Hartmann, Klebanoff) to intuitive, impressionistic interpretations (e.g., Bellack, Rotter and Jessor). Some of the analysts note regretfully that Shneidman did not present the entire TAT or MAPS test for their analysis, thereby in some cases eliminating pictures which the individual analyst considered important.

And what of the "supplementary tests" and the psychiatric data which are included presumably for prediction checks? Three of the tests were given before insulin therapy,

two after it. Moreover, how are these tests which to be used? If they are to serve as validation data to evaluate the relative merits of the TAT and MAPS are we ready to accept the Rorschach or the Draw-A-Person, for example, as a validating instrument for the TAT or MAPS in the same way that we accept the Wechsler-Bellevue IQ as a measure against which to test estimates of John Doe's intellectual capacities as revealed by the TAT? These are questions which clinicians cannot answer without very precise experimental study.

On the other hand, if the psychiatric material is to be used as the validation data, here, too, Shneidman's material is poor. Five different psychiatrically trained workers interviewed John Doe from time to time. Some were social workers, others were medical men. The psychotherapy undertaken with John Doe was apparently sporadic. He seems to be one of those patients who wander in and out of hospitals and outpatient clinics, who are seen by many people for longer or shorter periods of time, and for whom no systematic treatment plan evolves.

How seriously do these methodological factors limit the usefulness of these materials? Shneidman acknowledges that they do constitute "a limitation upon the possible conclusions" (p. 4). "This investigation," he continues, "is a research but not an experiment; it employed systematic observation, not controlled experiment." This admission hardly excuses the lack of scientific rigor. To be sure, clinical settings often do set limits on a research design. John Doe often cannot be manipulated. Yet test material on John Doe accumulates in the files because of the diagnostic demands of the hospital setting. The mere availability of case material, however, does not in itself justify its extensive manipulation for publication.

One final approach to this material remains for consideration. What does it reveal about the present state of thematic tests? Taking these fifteen scoring techniques as representative of the manner in which the TAT is now being used, it is clear that much has been added to the original need-press scoring scheme of Murray and Sanford even by those who retain some of the basic structure (e.g., Aron and Holt). Yet the fundamental problem of the meaning of "projection" has not been resolved. What level of personality do these tests tap? How much effect does the stimulus (TAT picture, MAPS setting, inkblot) have on what is revealed? Are these tests primarily useful as diagnostic tools or are they better for revealing underlying dynamics? The experts do not yet agree. Most of the analysts end up with the diagnosis of schizophrenia for John Doe. Some are of the opinion that the MAPS test shows a healthier pattern than does the TAT (Klebanoff, p. 131), others find even clearer evidence of pathology in the MAPS test (Arnold, p. 36; Korchin, p. 142). Some regard the MAPS test as better for diagnostic purposes, the TAT for the dynamics. Some stress the absence of guilt (White, p. 198), others find striking evidence of guilt (Bellak, p. 51). Since all of the analysts have had considerable experience as clinicians, the differences between them stem, presumably, from differences in scoring rationale. To resolve these differences research is indeed needed. Given the present materials the reader could choose between these schemes largely in terms of personal preference for the

normative vs. the intuitive approach, for example, or for present ego-functioning vs. an approach which emphasizes reconstruction of the past. He could not choose in terms of which is the most valid technique.

THELMA G. ALPER.

Wellesley College.

GELDARD, FRANK A. *The human senses.* New York: Wiley, 1953. Pp. x+365. $5.00.

This is a general textbook of the senses—a textbook of human sensory psychophysiology in the complete sense of these words. Its 323 pages of actual text are divided into 15 chapters. The first, 13 pages in length, is introductory. The next four, totaling 80 pages, are concerned with the sense of sight (The Visual Stimulus and the Eye; Basic Visual Phenomena; Color Vision and Color Blindness; and Visual Acuity, Contrast, and Interaction). Hearing is discussed in three chapters (Sound Energy and the Ear, Auditory Phenomena, and Electrophysiology and Auditory Theory), occupying 64 pages. The skin senses are allocated about as much space as vision and hearing, three chapters totaling 75 pages (The Skin and Its Stimuli, Pressure and Pain, Temperature Sensitivity). Kinesthetic and Organic Sensibilities, Labyrinthine Sensitivity, The Sense of Smell, and The Sense of Taste, are the headings for the last four chapters, which contain 16, 21, 25, and 29 pages, respectively. The text is followed by an excellent list of 330 references and an unusually detailed index for a book of this size. The latter is 25 pages long and must contain well over 2,000 entries.

Perhaps the most striking thing apparent in the statistical description above is the relative amount of space allocated to vision and hearing as compared with the other senses. Geldard's treatment of the former occupies a scant third of the entire book. Boring, Langfeld, and Weld (*Foundations of Psychology*), Munn (*Psychology*), and Stevens (*Handbook of Experimental Psychology*), on the other hand, assign vision and hearing the greatest prominence; in all three cases these two senses are allocated almost exactly 75 per cent of the space devoted to sensory material generally. Geldard's book, though unbalanced, represents a reasonable distribution of emphasis since there are several good textbooks concerned exclusively with vision and hearing, and scarcely any with the other senses.

As regards the general level of his writing, Geldard lies between Boring, Langfeld, and Weld, or Munn, on the one hand, and Stevens, on the other. *The Human Senses* contains considerably more advanced and detailed material on the senses than one finds in the typical good introductory text, but it is much less exhaustive and scholarly than Stevens' work. Taking Boring *et al.* and Stevens as the two ends of a scale, I should say that Geldard lies closer to the former than the latter. In fact, the chances are that the experimental psychologist who has kept reasonably abreast of recent developments will find Geldard's treatment a little too elementary for his taste.

Geldard's exposition is lucid throughout and his position is generally eclectic as regards sensory theories. I found no glaring errors to comment upon, or outrageous points of view to quarrel with. My chief complaint is rather that he committed too many errors of omission. However, if you take this book for what it is—an intermediate-level

text with a sound, general perspective—you will have to agree that it fills a definite gap in the psychologist's library.

A. CHAPANIS.
Bell Telephone Laboratories
Murray Hill, New Jersey.

NORTHWAY, MARY L. *A primer of sociometry.* Toronto: Univer. of Toronto Press, 1952. Pp. vi+48. $2.25, cloth-bound; $1.50, paper-bound.

A sociometric test measures the spontaneous choices of the members (or would-be members) of a group for and against associating with one another. Obviously, it must therefore employ an appropriate criterion for the choosing. Thus, in work groups, this would be "to be in the same work unit," "as neighboring workers on an assembly line"; in hospitalized groups, this could be "to be on the same ward," etc. Once the patterns of interpersonal response are uncovered, the meaning of the choices for the individual himself, for the functioning of the group, for the importance they have in the areas of social and clinical psychology, sociology, and anthropology, are the concern of sociometry—namely, under what conditions of group life and the personal past life of individual members will such and such patterns develop. The intricacies of the patterns and the complexities of their theoretical implications are enough to make many a psychologist hesitate to pursue sociometric research persistently. It is to all workers, as well as to beginners, in this position that Northway's *A Primer of Sociometry* is addressed, and is required reading.

The book orients the beginner to the nature and role of sociometry and provides the advanced worker with perspective of what has been accomplished and what lies ahead as untouched territory for investigation. It achieves these ends in 48 pages of refreshing, clear, and compact writing. The need for such a publication is widely felt and this one bids fair to become a classic of its kind in the midst of a literature which over the last 25 years contains about as many unreliable as dependable reports.

Well-balanced and critical discussion is directed upon the merits and limitations of sociometric method, analysis, and evaluation of results. To this reviewer, the only unimportant or irrelevant discussion for the purposes of the presentation appears to be the brief use of comparisons with other methods. Practically every major aspect of sociometric research (with the omission of negative choice or rejection) is treated, followed by specific references to the most authoritative works bearing upon that aspect. Many varieties of Canadian, French, and American work are described in context under the particular aspect they shed light upon. The presentation is organized in relation to theory, practice, and application in research design. Problems are outlined which raise our sights on research into what is growth and what is mental health.

Last, not least, the author gives profoundly suggestive treatment to elements which appear likely to remain universally important to choice reactions between people. Study directed toward them may be fruitful for many generations. Hence, *A Primer of Sociometry* can be seen as a practically useful and theoretically significant research guide to students of social, emotional, and group processes.

HELEN HALL JENNINGS.
Brooklyn College.

WEINLAND, JAMES D. AND GROSS, MARGARET V. *Personnel interviewing.* New York: Ronald, 1952. Pp. vii+416. $6.00.

Personnel Interviewing by Weinland and Gross is not likely to receive much attention from psychologists. The authors have not approached their goal of having the book ". . . serve as a guide for all those who are concerned with the selection of personnel and the maintenance of sound personnel relations" (p. iii). This text seems to be aimed at individuals in the personnel field who have had no training in psychology and have no intention of getting such training.

The authors have taken upon themselves all-inclusive and ambitious purposes: "In recent years, scientific research has made available to supervisors of interviewers and personnel managers many new techniques and applications of personnel interviewing. Some of these, such as the non-directive, group, board and stress interviews and the Chappel [*sic*]-Chronograph have been used successfully in industrial personnel work. A major purpose of this book is to explain these practical and valuable developments within a comprehensive discussion of interviewing methods, and to show how they can be employed to advantage in firms of all sizes" (p. iii). The extent to which the authors achieve their purposes is illustrated by the fact that the entire space devoted to a discussion of the Chapple Chronograph consists of two short paragraphs, and one can hardly hope to learn much about the Chapple Interaction Chronograph from this description (p. 113).

Some statements in the book cause one to wonder upon which general psychological principles the book claims to be based: e.g., "Certain wrinkles in older people tell of repeated grimaces; others of repeated smiles. The interviewer who is interested in people and takes care to study them can often tell a good deal about a person just by looking at him" (p. 16); and "Lavish jewelry suggests an ego-centered personality . . ." (p. 272).

The authors have attempted in this one short volume to deal with individual differences, personality dynamics, motivation, directive and nondirective interviewing, tests, types of interviews, correlation, etc. The shortcomings of their treatment of the various topics is to be expected in the light of their ambitious attempt. The result will probably leave the uninformed reader in a state of confusion. For example, on page 242 the authors state: "A correlation of '0.87 ± 0.12' would indicate that in a particular experiment a correlation of 0.87 was found but that the same correlation might not be found if the same tests, or measures, were used in another case. Correlations must always be taken with a 'grain of salt,' the size of the probable error showing how much 'salt' to use." These statements must leave the reader still not knowing how much "salt" is to be applied to the correlation coefficient of "0.87 ± 0.12."

A book on personnel interviewing is needed in the industrial field. It is doubted by this reviewer that the contribution by Weinland and Gross begins to fill the need.

M. J. WANTMAN.
University of Rochester.

ARBOUS, A. G. *Tables for aptitude testers.* South Africa: National Institute for Personnel Research, 1952. Pp. 3+86.

BLITSTEN, DOROTHY R. *The social theories of Harry Stack Sullivan.* New York: William-Frederick, 1953. Pp. 186. $3.50.

BOSSELMAN, BEULAH C. *The troubled mind; a psychiatric study of success and failure in human adaptation.* New York: Ronald, 1953. Pp. iv+206. $3.50.

BROWN, J. S., HARLOW, HARRY F., POSTMAN, LEO J., NOWLIS, V., NEWCOMB, T. H., & MOWRER, O. HOBART. *Current theory and research in motivation; a symposium.* Lincoln: Univer. of Nebraska Press, 1953. Pp. v+193. $2.00.

BROSS, IRWIN D. J. *Design for decision.* New York: Macmillan, 1953. Pp. viii+276. $4.25.

CARROLL, JOHN B. *The study of language.* Cambridge: Harvard Univer. Press, 1953. Pp. xi+289. $4.75.

DANIEL, ROBERT S., & LOUTTIT, C. M. *Professional problems in psychology.* New York: Prentice-Hall, 1953. Pp. xv+416. $5.50.

DOLLARD, JOHN, AULD, FRANK, JR., & WHITE, ALICE MARSDEN. *Steps in psychotherapy; study of a case of sex-fear conflict.* New York: Macmillan, 1953. Pp. ix+222. $3.50.

EYSENCK, H. J. *The structure of human personality.* New York: Wiley, 1953. Pp. xix+348. $5.75.

FESTINGER, LEON, & KATZ, DANIEL. (Eds.) *Research methods in the behavioral sciences.* New York: Dryden, 1953. Pp. xi+660. $5.90.

HALL, VICTOR E., CRISMON, JEFFERSON M., GIESE, ARTHUR C. (Eds.)

Annual review of physiology. Stanford: Annual Reviews, 1953. Pp. ix+558. $6.00.

HAVIGHURST, ROBERT J., & ALBRECHT, RUTH. *Older people.* New York: Longmans, Green, 1953. Pp. xvi+415. $5.00.

HURLOCK, ELIZABETH B. *Developmental psychology.* New York: McGraw-Hill, 1953. Pp. ix+556. $6.00.

JONES, ERNEST. *The life and work of Sigmund Freud.* Vol. I. New York: Basic Books, 1953. Pp. xiv+428. $6.75.

KATTSOFF, LOUIS O. *The design of human behavior.* St. Louis: Educational Publishers, 1947. Pp. x+402. $5.00.

KINSEY, ALFRED C., POMEROY, WARDELL B., MARTIN, CLYDE E., & GEBHARD, PAUL H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953. Pp. xxx+842. $8.00.

LAUCKS, IRVING F. *A speculation in reality.* New York: Philosophical Library, 1953. Pp. 154. $3.75.

LEHMAN, HARVEY C. *Age and achievement.* Princeton: Princeton Univer. Press for the American Philosophical Society, 1953. Pp. xi+359. $7.50.

LEITNER, KONRADI. *Hypnotism for professionals.* New York: Stravon, 1953. Pp. 127. $4.00.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953. Pp. xix+393. $6.50.

LITTLE, WILSON, & CHAPMAN, A. L. *Developmental guidance in secondary school.* New York: McGraw-Hill, 1953. Pp. xi+324. $4.50.

MANNHEIM, KARL. *Essays on sociology and social psychology.* (Paul Kecskemeti, Ed.) New York: Oxford Univer. Press, 1953. Pp. viii+319. $6.50.

MORSE, NANCY C. *Satisfactions in the white-collar job.* Ann Arbor: Univer. of Michigan Press, 1953. Pp. viii+235. $3.50.

MOWRER, O. HOBART. *Psychotherapy; theory and research.* New York: Ronald, 1953. Pp. xviii+700. $10.00.

OSGOOD, CHARLES E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953. Pp. vi+800. $10.00, college edition. $14.00, trade edition.

OVERHOLSER, WINFRED. *The psychiatrist and the law.* New York: Harcourt, Brace, 1953. Pp. x+147. $3.50.

PECK, LEIGH. *Child psychology: a dynamic approach.* Boston: Heath, 1953. Pp. 536. $5.25.

PERRY, HELEN S., & GAWEL, MARY L. (Eds.) *Harry Stack Sullivan: the interpersonal theory of psychiatry.* New York: Norton, 1953. Pp. xviii+393. $5.00.

PSYCHOTHERAPY RESEARCH GROUP, PENNSYLVANIA STATE COLLEGE, Wm. U. Snyder, Chairman. *Group report of a program of research in psychotherapy.* State College: Pennsylvania State Coll. Press, 1953. Pp. iii+179. $2.25.

RALLI, ELAINE P. (Ed.) *Adrenal cortex.* New York: Josiah Macy, Jr. Foundation, 1953. Pp. 165. $4.00.

REDFIELD, CHARLES E. *Communication in management.* Chicago: Univer. of Chicago Press, 1953. Pp. xvi+290. $3.75.

REIK, THEODOR. *The haunting melody.* New York: Farrar, Straus and Young, 1953. Pp. viii+376. $4.00.

SARGENT, HELEN D. *The insight test; a verbal projective test for personality study.* New York: Grune & Stratton, 1953. Pp. xii+276. $6.75.

SEGAL, J. *Le mecanisme de la vision des couleurs.* Paris: G. Doin & Cie, 1953. Pp. 351. 3.000 fr.

SEREG, MAX. *New light on dreams.* Boston: House of Edinboro, 1953. Pp. 159. $3.00.

SHERIF, MUZAFER, & WILSON, M. O. (Eds.) *Group relations at the crossroads.* (The University of Oklahoma lectures in social psychology.) New York: Harper, 1953. Pp. viii+379. $4.00.

SLATER, ELIOT, & SHIELDS, JAMES. *Psychotic and neurotic illnesses in twins.* London: Her Majesty's Stationery Office, 1953. Pp. v+385. $4.75.

SMUL, JOSEF S. *Respiratory diseases and allergy.* New York: Medical Library, 1953. Pp. 80. $2.75.

STEPHENSON, WILLIAM. *The study of behavior; Q-technique and its methodology.* Chicago: Univer. of Chicago Press, 1953. Pp. ix+376.

TYLER, LEONA E. *The work of the counselor.* New York: Appleton-Century-Crofts, 1953. Pp. xi+323. $3.00.

VITELES, MORRIS S. *Motivation and morale in industry.* New York: Norton, 1953. Pp. xvi+510. $9.50.

WALKER, HELEN M. *Statistical inference.* New York: Holt, 1953. Pp. xi+510. $6.25.

WORTIS, JOSEPH. (Ed.) *Basic problems in psychiatry.* New York: Grune & Stratton, 1953. Pp. 186. $4.50.

YOST, ORIN ROSS. *What you should know about mental illness.* New York: Exposition Press, 1953. Pp. x+165. $3.50.

Psychological Bulletin

A RECONSIDERATION OF THE PROBLEM OF INTROSPECTION

DAVID BAKAN
University of Missouri

It is the purpose of this essay to raise a general question for rethinking in the perspective of modern times. Two related considerations are involved in the motivation to write and publish an essay on introspection as a method for the investigation of psychological phenomena. The first is a sense of society's need for a psychology which is more appropriate to its problems. The second is a conviction that although psychologists should be methodologically careful, they should not afford themselves the luxury of methodological snobbery. There is no investigatory method which is "pure," and which provides an absolute guarantee against the commission of error. If errors be committed, we look to the future for their correctives. In the meantime, and perhaps ultimately, we accept a pragmatic criterion.

It is characteristic in the history of ideas that when some notion is rejected, even for adequate cause, many seemingly associated notions get rejected with it. Often these associated notions may be sound. One of the theses of this essay is that such has been the case with introspection. In the outright rejection of the method of introspection, much that was of considerable value was rejected.

In spite of the avowed rejection of the method, it has stayed with us in several disguised forms. As Boring has recently indicated, "introspection is still with us, doing its business under various aliases, of which *verbal report* is one" (4, p. 169). Boring seems relatively uncritical of the manner in which we contemporarily avail ourselves of introspection. The argument here is for a careful and avowed use of introspection.

In less disguised form introspection is with us in contemporary clinical psychology. The method of introspection is the method that the patient uses, although there is little avowed recognition of it as method of the clinician, except perhaps among the psychoanalysts (15). However, "therapy" is coming to be viewed as appropriate training for the aspirant clinician even in non-psychoanalytic contexts.

A HYPOTHESIS CONCERNING THE REJECTION OF THE INTROSPECTIVE METHOD

The rejection of the method of introspection is coincident with the inception of behaviorism in America. The first important behavioristic pronunciamento took place in 1913 (20). It is important to understand the immediate antecedents of behaviorism in order to understand the wide popularity that it gained. Boring's comprehensive history makes it unnecessary to recount the involved circumstances associated with the death of classical introspection.

Boring believes that it "went out of style . . . because it had demonstrated no functional use and therefore seemed dull, and also because it was unreliable" (4, p. 174). In the next few paragraphs a hypothesis will be offered to supplement that of Boring.

Psychology was in the throes of the Wurzburg-Cornell struggle in the first decade of the twentieth century. The Wurzburgers had discovered imageless thoughts; and they themselves hardly knew what to do with them. Titchener, at Cornell, sensed the staggering implications of the Würzburg findings, and struggled desperately to reject them (16).

The psychological literature of the time is in many respects confused, repetitive, and—we might say—anguished.   Psychology had, it seemed, got itself into absolutely inextricable difficulties; and there was no one within the introspective movement who had the clarity of vision to go beyond these difficulties. Watson, for all the limitations that we may ascribe to him, had clarity and offered a program which psychologists could follow.

Let us briefly examine the nature of some of the Würzburg findings. They discovered that thought was possible without images; and that thought was guided by states variously designated by the terms *Aufgabe, Bewusstseinslage,* and *determinierende Tendenz.* The favored method was the *Ausfragemethode.* Mayer and Orth (11) used the method of free association to a verbal stimulus, instructing the subject to report *everything* that went on between the hearing of the stimulus word and the making of the response. Messer (14) finds himself forced to posit *unconscious* processes underlying the processes of thought. Ach (1, 9)[1] introduces the concept of the will, i e., motivation, as guiding the thought processes; he uses a probing investigatory procedure; and he uses hypnosis. Bühler (6) indicates that it is important, in the study of the thought processes, to empathize and sympathize with the subjects engaged in this kind of experimentation.

Then, the problem is dropped like the proverbial hot potato.   Külpe, the leading figure in the Würzburg movement, leaves Würzburg and goes to Bonn in 1909, and the work practically ceases.   Bühler posthumously publishes Külpe's lectures which, according to Boring, "contain a pretty complete system of psychology. But the chapter on thought was missing! Bühler said that Külpe had not been lecturing on the topic" (3, p. 407).

In the light of the foregoing, and in the light of what we have learned from psychoanalysis, a rather simple explanation suggests itself. These investigators were using themselves and each other as subjects. They had struck the unconscious, and particularly unconscious motivation, and had to probe it if they were to make any headway. However, as we know today, probing the unconscious tends to generate anxiety and resistance; and these investigators simply were not prepared to undergo the necessary personal trials involved. Boring (4, p. 186) suggests a relationship between the Würzburg school and Freud, but makes little of it.

Psychology had two possible alternatives: either to widen its investigations to take account of and to study the role of unconscious motivation on the thought processes, or to detour.   Academic psychology detoured; and detoured in two ways:

---

[1] The writer could not locate a copy of Ach's book.   This sentence is based on Humphrey's (9) summary of Ach's work.

It detoured by way of behaviorism, completely rejecting (at least avowedly) the whole method of introspection, and it detoured by way of gestalt psychology. The former dropped the whole concept of mind, conscious and unconscious. The latter adopted as a basic principle that whatever introspection is done should be *naive* introspection, with no probing and no analysis, thus presenting intrusion upon the unconscious.

A BASIC DISTINCTION FOR INTROSPECTION

Perhaps one of the most important distinctions necessary for the understanding of the nature of introspection is the classical one between the experience and that of which the experience is. It is the distinction which is contained in the classical one of *Kundgabe* versus *Beschreibung* (4). It is the distinction which is indicated by the concept of stimulus-error (2). It is the distinction which the psychoanalyst makes when he concerns himself primarily with a memory, as contrasted with the event to which the memory presumably refers.

The distinction is somewhat difficult to grasp when we deal with perception. Let us consider a simple experience reported as "I see a book." From the point of view of this distinction it is one or another of two reports: "I *see* a book," or "I see a *book*." In the first instance it is a report of experience as experience. In the second instance the reference is to the object rather than to the experience of the object. The first can be true, and the second false, as, for example, in an hallucination.

The distinction is easier to make when we consider something like anxiety. It is hard to make when the experience involves an external stimulus. It is of interest that when Washburn made her presidential address before the American Psychological Association (19) in 1921 she felt that it was necessary to say that introspection is proper only where there is an external stimulus. This, she believed, would endow introspection with "objectivity"—an unfortunate semantic identification of "object" with "objectivity." It is here, probably, when the Watsonian noose was drawing very tight around the neck of introspection, that introspection surrendered the very thing which was its major merit. Introspection has its maximum value on those very experiences for which there may be no conspicuous physical stimuli, such as grief, joy, anxiety, anger, depression, exhilaration, etc.

THE PROBLEM OF LANGUAGE AND COMMUNICATION

A major criticism which has been leveled against the method of introspection is that the *data* of introspection are not public. In the case of overt behavior it is possible, at least in principle, for two observers to observe a given phenomenon simultaneously. This has sometimes been referred to as the criterion of publicity; and it has been said that data are not acceptable unless this criterion has been satisfied (again, at least in principle).

That introspective data are not public in this sense is not to be questioned. What is to be questioned is whether the criterion is essential. What is the value of the criterion of publicity? Its value, presumably, inheres in the conviction that it avoids error and provides for verification. However, can we not have verification without publicity? Let us consider one of the most acceptable kinds of investigatory procedure from

this point of view, the conditioning experiment. There is no way of verifying Pavlov's experiments today by having another observer watching them, since, to say the least, Pavlov's dogs are probably all quite dead. In order to verify Pavlov's findings we would have to get other dogs. Furthermore, the fact that two people could have stood by to count the number of drops of saliva is quite irrelevant. If the criterion of publicity is not met by introspection, it is not really very serious as long as each scientist has, so to speak, at least one "dog" whom he can observe directly.

The crisis which was generated by disparate results from Würzburg and Cornell, with the one finding imageless thoughts and the other not finding them, was hardly adequate reason for the total rejection of introspection. Disparate results from different laboratories are usually provocative of further investigation, rather than the occasion for dropping the problems, the methods, and the fundamental points of view involved. The failure of the introspective method to satisfy this naive criterion of publicity could hardly have been the real reason for the rejection of introspection as a method.

A more important problem is the possibility of publicity, not of the data, but of the report. Even though the process of introspective observation is, in a sense, private, the information gleaned from the observations must be public. This raises the question of language and communication. There are two questions that may be asked in connection with language with respect to introspection: First, if we relate our introspections to one another, would we understand one another? Second, if we do understand one another, how does this come to pass? If the answer to the first question is to any degree affirmative, then to that extent is the criterion of publicity of report satisfied.

For the answer to the first question we appeal, at the very least, to common sense. If we hear a person say, "I am sorry," or "I am worried," or "I feel sick," etc., there is hardly any question but that we understand what he means. There are times when we may not *believe* him; but the possibility of fraud, intentional or unintentional, or of lack of precision exists with respect to any methodology. The fact is, however, that we understand him.

The answer to the second question now becomes a matter for empirical investigation. This is not the place to enter into a detailed discussion of the psychology of language learning. However, it is extremely pertinent to indicate that the theory of language learning implicit in contemporary behavioristics is much more simple than is consistent with the facts. This implicit theory may be roughly characterized as follows:

The teacher holds up a ball and says, "Ball." The learner repeats, "Ball." The learner then, presumably, comes to "know" the meaning of the word. Certainly the theory is stretched to the breaking point when presented with the fact that we all fairly well understand the meanings of words such as "sorrow," "feeling," "nausea," "if," "but," etc.

INTROSPECTION AS RETROSPECTION

In 1921 Titchener (18) wrote an essay which, in part, attempted to present to English-speaking readers some of the contributions of Franz Brentano. In the judgment of the writer, Brentano is one of the most important figures in the history of psychology. The major work of Brentano with respect to psychology (5), has not, as far as could be determined by the writer, been translated

into English. Of Brentano and Wundt, Titchener wrote: "The student of psychology, though his personal indebtedness be also twofold, must still make his choice for one or the other. There is no middle way between Brentano and Wundt" (18, p. 108). For the most part, the choice of the classical introspectionists was for Wundt. Wundt and Brentano published their major psychological works at about the same time. Two major schools of thought issue from Brentano. One is the already-mentioned Würzburg school. The other is psychoanalysis, with Brentano having been the only academic psychologist under whom Freud studied (12, 13). Psychoanalysis, however, differed from the Würzburgers with respect to a readiness to face the unconscious. It may have been easier for Freud to break through to the unconscious because it was not his own unconscious but the unconscious of his patients. It was only secondarily that Freud used himself as subject. The Würzburgers, on the other hand, used themselves and each other as subjects.

Brentano, Külpe, and Freud conceived of introspection not as of the present, but the past. They took seriously what was then a common observation that introspection at the moment an experience is taking place changes the character of that experience. If we are interested, say, in anger, then introspection at the moment of anger tends to reduce the anger. It is only when anger is past that it can be properly examined. Using the method of introspection, thus avowedly retrospectively, makes it possible to examine psychological phenomena which cannot readily be elicited in the laboratory, except perhaps with very great ingenuity.

This difficulty of the introspection of Wundt and Titchener was adequately recognized by McDougall. He wrote: "Experimental introspection has obvious limitations. Many of our most vital and interesting experiences, such as grief or joy or fear or moral struggle, cannot be induced at will, except perhaps, in very slight degrees. And, under the most favourable conditions, introspection of our more vivid and vital experiences is difficult, because we are apt to be primarily interested in the events of the outer world in which we are taking part, if only as observers. Then again the very act of introspection does to some extent modify the experiences we wish to observe and describe; so that in introspecting we partially defeat our own purposes" (10, p. 4).

Thus, the type of introspection which was advocated by Titchener, and which was the object of attack by the anti-introspectionists, was a type which, by its nature, could not attack the important aspects and kinds of experience. The cry that a psychology was wanted which would have some usefulness was completely justified when the object of attack was the kind of introspection advocated by Titchener.

Errors of Introspection

A characteristic of good science is that it is ever alert to the possibility of the commission of systematic types of errors. One of the major criticisms which has been leveled against introspection is that its results are untrustworthy. In the following few paragraphs a brief attempt will be made to examine the problem of the trustworthiness or validity of introspective reports.

There is a respect in which introspective observations are more trustworthy than observations made by the use of the sense organs. Sense organs may be defective. Sense or-

gans are subject to illusion. Observations made with the sense organs are subject to the accidents of angle of regard, kind of illumination, noise level, etc. In the last analysis, the sense organs are subject to hallucination. Introspection is a method which does not involve the sense organs in the usual fashion, and therefore all of the error tendencies associated with the sense organs do not exist for introspection.

However, introspection has associated with it other sources of error. But even at this date, we have achieved a certain amount of progress in isolating them. We know about the stimulus-error. We are aware of the tendency to suppress data (repression), of the tendency to supply socially acceptable data in place of other data (distortion, rationalization, displacement, etc.). But, insofar as we are aware of these error tendencies we can take precautions against their commission. In this respect introspection is no different from any other set of methods in science. To be aware, for example, of the tendency toward rationalization stimulates us to challenge our introspective findings to determine whether they have resulted from the rationalization process. It is a matter of time and careful work to discover other error sources. We have discovered suggestion, cultural determination, ethnocentrism, etc.; and the list will probably lengthen as our experience with the method is enlarged.

A FUNDAMENTAL DIFFERENCE BETWEEN CLASSICAL INTROSPECTION AND PSYCHOANALYSIS

Psychoanalysis has one major limitation with respect to our purposes which was not present in classical introspection. This is that the major objective of psychoanalysis is therapy.[1] The major objective of the classical introspectionists was the acquisition of knowledge. This is a fundamental difference.

Essentially, what is being advocated in this paper is the use of the psychoanalytic *method* with the *objective* of the classical introspectionists.

It has been indicated that what is being advocated in this paper is partly on the grounds of the need for a science of psychology which has practical implications. However, there is an old lesson in the history of science of which we avail ourselves. Whereas knowledge may have practicality as its ultimate objective, it has been found that we sometimes do better, both practically and theoretically, if we temporarily forsake the practical objective.

In taking the objective from the classical introspectionists, it is necessary to make some modification in the psychoanalytic procedure. Although the investigator should be "free" in his associations, he should not permit himself to wander too far from the subject under investigation. His associations should stay under the influence of the task at hand. Of course, as in any investigation, decisions of relevance have to be made, and sometimes only a dim intuition dictates the nature of these decisions. Although there is no a priori method for determining relevance, the investigator should always attempt to keep in mind that he is serving science primarily, and himself secondarily.

[1] This is true even though Freud did envisage that "the future will probably attribute far greater importance to psychoanalysis as the science of the unconscious than as a therapeutic procedure" (7, p. 673).

A "MINIATURE" INVESTIGATION OF THE RETENTION AND REVELATION OF SECRETS BY THE METHOD OF RETROSPECTIVE ANALYSIS

In accordance with what has already been said the writer attempted to conduct an investigation of the kind suggested. It is a "miniature" investigation in that it was conducted only over a very short period of time. It was conducted for five days for about an hour and a half each day.

There were several reasons for the choice of the topic, retention and revelation of secrets. One of these is that the topic seemed to be one which is more amenable to introspective investigation than to other methods. By its very nature a secret is something which may not reflect itself in overt behavior. The latter almost constitutes a definition of a secret. Another reason for the choice of the topic is that it seems to be a fundamental one for any kind of introspective investigation. It seemed important to obtain information concerning the nature of secret retention and secret revelation before very much progress could be made with other topics. A third reason was that the topic seemed to lie close to the oft-stated objective of psychology as being prediction and control of human behavior.

The procedure simply involved sitting down to the typewriter and typing whatever came, after the decision concerning the topic was made. The choice of the typewriter was made primarily on the basis that the writer has found himself to be more fluent this way than either writing by hand or talking into a recording machine.

By virtue of the nature of the subject chosen, the writer attempted to write "as though" the material would never be released. Under any circumstances, even if this was a myth, the sense of the possibility of editing was not mythical. At the moment the writer does not consider it wise to release the protocol. However, one example will be given. The following is taken from the record with some editing:

... What is one of the secrets such as *thee and me* have? I once talked to a professor of zoology at lunch about the academic life. He commented that over the head of every academician hangs a sword on a thin string. No matter how much you do, you never feel that you are doing enough. I am reminded of Freud's dream of Irma's injection. He says, "I am always careful, of course, to see that the syringe is perfectly clean. *For I am conscientious.*" The italics are mine. If he felt that he were really conscientious, if he had no feelings of shortcomings in this connection, why did he have to protest that he was conscientious? The guilt of lack of conscientiousness haunts most of my friends. My lack of conscientiousness is my "secret." But here I find myself confessing to lack of conscientiousness. But I was not able to do so until I was able to remember something which would make it possible for me not to have my guilt alone. I brought up the zoology professor. When I wrote the above line about him I hesitated for a moment on the question of whether or not to use quotation marks, or to write it in the way that I did. The quotation marks would have had to come, in all honesty, after the word "string." I wrote on, however, "No matter how much you do, you never feel that you are doing enough." This is what I would have liked him to have said. I added it to give the impression that he had said it, but not quite lying about it.

I think that what has been said above can be generalized. *We are more prone to confess a secret guilt when we can believe that others have the same secret guilt . . . .*

The general pattern involved in this kind of writing is that of an oscillation between a free expressive mood and an analytic mood, with the free expression being the subject of the analysis. The question of what a given item of free expression might mean with respect to the major topic

under investigation was repeatedly asked.

In the course of this investigation a series of propositions, including the italicized one above, were formulated. This list can be considered to be the *yield* of this "miniature" investigation:

1. (Given above.)

2. Persons with a secret guilt tend to create situations in which they can "see" that others have the same secret guilt.

3. A secret is a secret by virtue of the anticipation of negative reactions from other people.

4. A secret is maintained in order to maintain some given perception of one's self in others.

5. Persons who associate with one another in the context of a larger group, who have a secret from that larger group, will create a metaphorical or otherwise cryptographic language in which to discuss the secret.

6. In order to conceal a secret one may tend to reveal a fabricated "secret," or a less-secret secret, in order to generate the impression that one is being open and frank.

7. One of the important secret areas in our culture is in connection with our intellectual limitations.

8. When an individual has a secret he will attempt to "protest" that the opposite is the case, if the secret has an opposite.

9. The revelation of a secret may involve the attempt to generate the impression that one is telling a joke, to achieve the double purpose of revelation on the one hand and disbelief on the other.

10. In the revelation of a secret one may attempt to generate the impression that one degrades one's self in one's own eyes, in order to reduce the degradation that one anticipates will be the reaction of others to the revelation.

11. If A knows a secret about B and B knows a secret about A, and if A discovers that B has revealed A's secret, then A will be inclined to reveal B's secret.

12. If an individual changes his group identification from Group A to Group B, and if Group A has a secret which it keeps from Group B, that individual will be inclined to reveal Group A's secret to the members of Group B.

DISCUSSION OF THE "MINIATURE" INVESTIGATION

The simple fecundity of the method soon became evident. After the decision was made to attempt it and a brief beginning was made, it became apparent that this was, to use a term from the vernacular, a veritable mine of information. Essentially it capitalizes on the fact that the investigator has had twenty or thirty or forty or fifty or sixty or seventy years for the collection of various kinds of information. Certainly one of the defects of this kind of data collection is that it is not systematic in the usual way in which we understand this term. Yet it is the result of years of trial and error, of a kind which most laboratory types of investigation do not generally get. It may be argued that these data have been uncritically gathered. This is a valid point. However, the necessary criticality can be supplied in the course of the investigation itself.

This kind of investigation can be severely hampered by what may be loosely designated as "ethical" considerations. Let us consider, for example, the third proposition enumerated above, that a secret is maintained in order to maintain a given perception of one's self in others. By virtue of the intimate connection between ethics, in this larger sense, and the kind of data which may become

the subject of an introspective investigation, it is extremely important that the investigator attempt, to the degree that he can, to divest the investigation of ethical considerations. Methodologically this divesture may involve a preliminary investigation of the ethical considerations themselves. Also, it must be added that, for some kinds of problems to be investigated by these methods, less may be required in the way of preliminary investigation than for other problems. However, for the investigation of any problem by these methods, a scientific and objective attitude is a prerequisite.

One of the major merits of this kind of an approach is that it studies the phenomena of psychology *directly*, in a manner which is rarely the case in most psychological investigations. Actually, the kind of material which issues from an introspective investigation of the kind being advocated is *presupposed* in many other psychological investigations. Consider for the moment the "lie" scale of the Minnesota Multiphasic Personality Inventory. The test presumably "gets at" the kind of thing which has been investigated in the investigation on secrets cited above. However, the items of this scale were selected because they would presumably be answered negatively by persons who were trying to put themselves "in the most acceptable light socially" (8). This *presumes*, with little qualification, the content of the fourth proposition, as well as about 15 preconceptions concerning the meaning of social acceptability. (There are 15 items in the "lie" scale.)

Furthermore, had the makers of the MMPI critically examined the nature of secrets in the way in which it has been begun in the above investigation, they would have seen that there are other dynamics of lying, in addi-

tion to the one of which they did avail themselves. For example, proposition 6 indicates that a certain amount of truth-telling may simply be a device for "covering up" one or more other lies. It may well be that the operation of the dynamic indicated by proposition 6 acts to depress the "lie" score when lying is really taking place. A full awareness of the kind of thing that issues from such an investigation can greatly enhance the effectiveness even of pencil-and-paper tests.

From a more theoretical point of view, if we seriously accept the mission of psychology as being that of the prediction and control of human behavior, the psychology of secrets is an important link in the chain of psychological findings and theory. Investigators, no matter what they are investigating, must be cognizant, at the very least, of the possibility of dissemblance when they use human subjects. To predict and control an individual's behavior it is important to know, for example, his group-identifications, his objectives, his values, etc. Many of these items of information are secret. They may even be secret to the subject himself. And under any circumstances they are not items of information which will be revealed readily. Thus, until psychologists develop a rather full understanding of the dynamics of this phenomenon, the ignorance of this phenomenon will stand in the way of other investigations.

What has been said in the above paragraph would be considerably less cogent if the phenomenon of the secret played only a small role in connection with other phenomena. However, secrets play their most important role in those phenomena which are most vital. A psychology that seeks to understand these vital phenomena must have an appreciation of the

phenomena of secret retention and secret revelation. Whether we are interested in the problems of marriage, industrial management, leadership, prejudice, loyalty, delinquency, international affairs, politics, military strategy, litigation, business practices, economics, etc., the psychology of secrets is extremely pertinent. And, the psychology of secrets yields most effectively to the method which is being proposed.

It might be indicated that an adequate psychology of secrets would represent an extremely important contribution to our country in its present state. An adequate psychology of secrets would provide us with an insuperable advantage over our enemies and potential enemies. Concretely, for example, we could use the information about the psychology of secrets for enhancing the skill of the military interrogator in extracting information from prisoners of war; and we could teach our own soldiers about possible techniques that may be used against them in the attempt to extract information from them, so that they may be prepared for them. Also, an adequate psychology of secrets might make quite unnecessary some contemporary political investigatory practices which enjoy some popularity only in lieu of more scientific devices.

Although the psychology of secrets is perhaps a central and basic one associated with the method, investigations could and should be pursued with great profit on other problems. Thus, for example, problem solving and decision making can and should be investigated by the method of retrospective analysis. Investigations on status, power, anxiety, fear, aggression, aesthetic experience, learning, communication, memory, concept formation, perception, judgment, charity, loneliness, betrayal, etc. could and should be carried out to enhance our understanding of these phenomena.

THE PROBLEM OF VALIDITY OF THE FINDINGS FROM THE "MINIATURE" INVESTIGATION

Perhaps the critical question in the mind of the reader up to this point is that of the "validity" of the findings of an introspective investigation. The problem of validity has already been discussed previously, but somewhat abstractly. In this section, the problem will be dealt with somewhat more concretely, with the findings of the "miniature" investigation before us.

The propositions which issued from the "miniature" investigation are, at least, what may be considered to be "hypotheses" for investigation by other methods. Thus, at the very least, the method may be recommended as a device for systematically getting hypotheses as contrasted with, say, the casual reaching out for a pair of variables and hypothesizing a relationship between them.

Again, as has already been indicated, it may be used as a method whereby an investigator can bring his presuppositions concerning an investigation to *formulation;* where he can critically examine his presuppositions; and where he might be helped in conceiving of other presuppositions against which he can contrast the ones he is using. Or, the method could be used as a device whereby an investigator, having got some experimental results which he cannot understand, provokes his imagination to arrive at some kind of an explanation of his results. The *deliberate* and *avowed* adoption of the method would be extremely helpful in these respects.

However, the writer believes that the method warrants more than this. As has been indicated, the method has a *directness* which is not to be found in any other method of investigation of psychological phenomena. In any investigation each thing which lies between the phenomenon and the data is a source of error. These sources of error are minimized by the method which is being proposed. All errors such as failure of the subject to cooperate (e.g., rehearsal when instructed not to do so in studies on reminiscence), dissemblance, failure to comprehend instructions, refusal to believe the expressions of the investigator's avowed intentions, fear of hidden—or manifest—microphones, lack of skill on the part of the subject (e.g., fixating a point in a vision experiment), refusal to take a "naive" attitude (in gestalt experiments), the lack of control over human subjects (e.g., subjects in problem-solving studies already knowing the solutions to problems but not informing the investigator), subjects knowing the intention of the investigator (e.g., subjects knowing that the experimenter is interested in demonstrating a relation between frustration and aggression, and therefore concealing their felt aggression), etc. are minimized in this kind of investigation.

The propositions which were yielded by the "miniature" investigation also have a certain kind of self-evidence associated with them. They elicit the "of course" response. Some of the propositions may require further specification and further qualification. Nevertheless, they are in some sense obvious. It is the sense of self-evidence which is associated, perhaps, with the axioms of Euclidean geometry. The nature of self-evidence is, of course, an extremely

difficult problem and perhaps more properly falls in the province of the philosopher. Or, perhaps, self-evidence is a problem to be investigated by the very methods which are here proposed. However, whatever the ultimate nature of self-evidence may be, there is a sense in which the results of an introspective investigation are of this type.

Now, of course, the matter of self-evidence may be challenged by the question: Self-evident to whom? In one respect this as a valid question. But in another respect it is not. It is valid in that if we are to know that it is self-evident it must be self-evident to someone. However, when the mathematician uses the term *self-evident* he means something which is intrinsic to the proposition, rather than something dependent upon the reader or the hearer of the proposition. For the mathematician it is the self-evidence of the proposition which makes it possible for the person to see the self-evidence, rather than the reverse. It is this characteristic which is shared by introspective propositions.

As a matter of fact, some of the propositions which issued from the "miniature" investigation seem to partake of greater self-evidence than others. Thus, for example, proposition 4 seems to be quite self-evident, whereas proposition 5 seems to be somewhat less self-evident. And even the seeming self-evidence of proposition 4 may be quite culture bound. However, what has been reported is only an extremely limited investigation, only a beginning and only a sample. Nevertheless, what has been presented is enough to suggest the possibility of achieving the kind of self-evidence that has been indicated.

Two related, but distinguishable, problems are those of replication and

generality. Can such an investigation be replicated? The answer is affirmative, although the difficulties of replication should be recognized and account should be taken of them. If an investigator attempts to replicate his own investigation at another time, he will inevitably be under the influence of what he has already done. In replicating such an investigation, the very replication itself should come under the scrutiny of the investigator. He should challenge, for example, his personal identification with the results he has already obtained, and prepare himself for finding both novelty and contradiction with respect to his earlier investigation. If one investigator is interested in replicating the investigation of another investigator, he should carefully take into account the possibility of suggestion, or of his willingness to accept the results of the earlier investigator (particularly if the first investigator has prestige for the second investigator). He should take careful cognizance of possible motivation for showing the earlier investigator to be in error, etc. In some instances it may be extremely worth while to investigate some topic without reading the results of the earlier investigation until the completion of the second investigation, and the making of a comparison later on. Carefully controlled experimentation to determine possible effects of suggestion, for example, is extremely feasible.

The generality of the results of such an investigation is a somewhat more difficult problem, but it is a difficulty which is not unique to introspective investigation. One investigator's results can be compared with another investigator's results, so that the problem of uniqueness with respect to a single investigator is vitiated. However, one may ask, in the event of consistency of results among a group of investigators, may the findings not be unique to a group of persons all of whom are introspective investigators? There is no easy answer to this problem. However, we face the same problem in other investigations. May not the results of studies in rote learning be largely unique to college sophomores? May not the results of studies in, say, secondary reinforcement be unique to rats, or more particularly laboratory rats, or even more particularly white laboratory rats, or still more particularly tamed white laboratory rats, etc.? May not all findings concerning mental abnormality be unique to mentally abnormal persons *contacted* by investigators, and may not these very contacts be a major determinant of the findings?

The answer, of course, to each of these questions is contingent upon some decision concerning relevance, a decision that has to be made in connection with any investigation. Actually, the kind of investigation being advocated has an advantage in this respect over other kinds of investigation. For, in an introspective investigation the very decisions concerning relevance can come under the same scrutiny as the phenomena being investigated.

POSSIBILITY AS A FINDING

The argument concerning the validity of the findings from an introspective investigation thus far has been concerned with validity in the usual sense, i.e., the argument has been concerned with the truth or falsity of propositions which issue from an introspective investigation.

There is, however, a value to such propositions which is over and beyond their validity. This is their *possibility* rather than their truth or

falsity. The knowledge that a certain dynamic is *possible* enhances the sensitivity of the psychological observer. To make this point concrete, let us again consider the military interrogation situation. Suppose that the interrogator is interested in determining the contents of some supplies which have been moved in by the enemy. Now suppose that the prisoner being interrogated knows these contents but does not wish to reveal them. The prisoner *may* avail himself of the dynamic indicated by proposition 6 (the revelation of less secret secrets in order to generate the impression that he is being open and frank), and inform the interrogator at length about a great number of lesser secrets, but not the nature of the supplies. He may say, "I will tell you everything that I know, but I do not know what was in those trucks." An interrogator who was not aware of the *possibility* of proposition 6 might be lulled into believing the man. The interrogator might say to himself, "he is evidently telling all that he knows." On the other hand, an interrogator who is aware of the possibility of the action of the dynamic indicated by proposition 6 would be aware of the possibility of this kind of deception, and would be less likely to be "taken in."

Insofar as at least one person can contrive such a device for deception, then such a device is *possible*, and some other individual *may* have conceived of it and *may* be making use of it. The truth of the proposition in this respect becomes quite secondary. What is important, simply, is that someone thought of it; and if one person thought of it, other persons might think of it.

In this respect psychologists can make a major contribution to society not only by rendering to society

established *truths*, but also by rendering to society established *possibles* with respect to psychological dynamics. In the matter of prediction and control of human behavior, a knowledge of what an individual might *possibly* do, or *possibly* feel, or *possibly* think places us well on the way toward the achievement of our objective. Given a detailed knowledge concerning the possibles we can act in such a fashion as to discourage some from becoming actualities, and to encourage others into becoming actualities. The pragmatic usefulness of knowledge of possibles extends from the clinical situation to world affairs.

As has been suggested, these possibles may indeed turn out to be truths in the larger and more scientific sense. But even if they fail to meet the criteria for general scientific propositions, they have value in the sense indicated above.

SUMMARY

It is the purpose of this essay to raise the question of the appropriateness of the method of introspection for rethinking in the perspective of modern times. Some features of the history of introspection in the first decade of the twentieth century have been pointed to. The hypothesis is advanced that introspection was dropped because the classical introspectionists had come to a point where they would have had to probe the unconscious to make any progress. From psychoanalysis we have learned that probing the unconscious generates anxiety and resistance. The old distinction between an experience, and that of which the experience is, is again made. The problems of communication of introspective observations are discussed. It is claimed that introspection should

be retrospective, consistent with the position of Brentano, Külpe, and Freud. The problem of errors in introspection is discussed. The position is advanced that retrospective analysis should take its *objective* from the classical introspectionists and its *method* from the psychoanalysts, with some modification. A "miniature" investigation on the psychology of secret retention and secret revelation is described. The implications of the investigation both as an investigation in its own right and as an example of the method are discussed. The prob-

lem of the validity of the findings is discussed. The problems of replication and generality are discussed. The propositions which issue from such an investigation can be viewed as having three values: first, as hypotheses to be investigated by other methods, and otherwise to supplement other methods; second, as propositions which have a certain kind of self-evidence associated with them; third, as *possibles* which can give us real assistance in achieving the objectives of prediction and control of human behavior.

REFERENCES

1. ACH, N. *Über die Willenstätigkeit und das Denken.* Göttingen: Vandenhoeck und Ruprecht, 1905.
2. BORING, E. G. The stimulus-error. *Amer. J. Psychol.*, 1921, 32, 449–471.
3. BORING, E. G. *A history of experimental psychology.* (2nd Ed.) New York: Appleton-Century-Crofts, 1950.
4. BORING, E. G. A history of introspection. *Psychol. Bull.*, 1953, 50, 169–189.
5. BRENTANO, F. *Psychologie vom empirischen Standpunkte.* Leipzig: Duncker und Humblot, 1874.
6. BÜHLER, K. Tatsachen und Probleme zu einer Psychologie der Denkvorgänge: I. Ueber Gedanken. *Arch. ges. Psychol.*, 1907, 9, 297–365.
7. FREUD, S. Psychoanalysis: Freudian School. *Encyclopedia Britannica*, 14th Ed., v. 18, pp. 672–674.
8. HATHAWAY, S. R, & McKINLEY, J. *Minnesota Multiphasic Personality Inventory manual.* New York: Psychological Corp., 1951.
9. HUMPHREY, G. *Thinking.* New York: Wiley, 1951.
10. McDOUGALL, W. Prolegomena to psychology. *Psychol. Rev.*, 1922, 29, 1–43.
11. MAYER, A., & ORTH, J. Zur qualitätiven

Untersuchung der Association. *Z. Psychol. Physiol. Sinnesorg.*, 1901, 26, 1–13.
12. MERLAN, P. Brentano and Freud. *J. Hist. Ideas.*, 1945, 6, 375–377.
13. MERLAN, P. Brentano and Freud—a sequel. *J. Hist. Ideas.*, 1949, 10, 451.
14. MESSER, A. Experimentell-psychologische Untersuchung über das Denken. *Arch. ges. Psychol.*, 1906, 8, 1–224.
15. REIK, T. *Listening with the third ear.* New York: Farrar & Strauss, 1948.
16. TITCHENER, E. B. *Lectures on the experimental psychology of the thought-processes.* New York: Macmillan, 1909.
17. TITCHENER, E. B. Description vs. statement of meaning. *Amer. J. Psychol.*, 1912, 23, 165–182.
18. TITCHENER, E. B. Brentano and Wundt: empirical and experimental psychology. *Amer. J. Psychol.*, 1921, 32, 108–120.
19. WASHBURN, MARGARET F. Introspection as an objective method. *Psychol. Rev.*, 1922, 29, 89–112.
20. WATSON, J. B. Psychology as the behaviorist views it. *Psychol. Rev.*, 1913, 20, 158–177.

RATING SCALES AND CHECK LISTS FOR THE EVALUATION OF PSYCHOPATHOLOGY

MAURICE LORR

Veterans Administration, Washington, D. C.

The use of check lists, charts, and rating scales for the objective recording and later evaluation of change in the behavior and symptoms of psychiatric patients is not new. Devices such as the Phipps Psychiatric Clinic Behavior Chart (4) have been used for a half century on psychiatric wards to record patient change. Plant (17) reported a rating scheme for describing patient behavior on the ward in 1922. In 1933, Moore (14) published his chart and "Schema for the Quantitative Measurement of Abnormal Emotional Conditions" containing some 36 carefully constructed scales. The interest in mental health problems intensified by events in World War II resulted in a marked upswing in the development of procedures for the objective measurement of psychopathology and personality change. It is the purpose of this review to examine briefly those rating scales and check lists designed to describe psychiatric patients on the ward or in the interview, which have appeared during the past ten years.

SCALES FOR USE BY PSYCHIATRIC AIDES AND NURSES

Six scales suitable for use in mental hospitals by nurses and psychiatric aides have been reported. These are generally designed to secure (a) quantified descriptive reports of readily observable patient ward behavior, and (b) quantitative estimates of hospital adjustment. The first of these, the Gardner Behavior Chart (22), a rating scale developed by Wilcox out of work with psychotic patients, is designed to secure reports of easily observed patients' ward behavior from nurses and attendants. The 15 categories or scales used are: attention to personal appearance, sleep, appetite, sociability, activity control, noise disturbance control, temper control, combativeness control, care of property, self entertainment, cooperation in routine, work capacity, work initiative when alone, work initiative when closely supervised, and willingness to follow directions. Under each category five brief phrases characterize the grades of behavior in the scale. The rating grades are none, poor, fair, good, and extra good; they are weighted from 0 to 4. The total score consists of the sum of the 15 ratings received. The Behavior Chart has been found useful in the evaluation of change following prefrontal lobotomy (21).

The Fergus Falls Behavior Rating Sheet, prepared by Lucero and Meyer (10), was developed to record behavior of patients who are mute, unintelligible, hyperactive, or seclusive. Eleven aspects of behavior, such as work, response to meals, and response to patients, are rated by checking one of five descriptions. A value of 1 is given to the most deviant behavior and a 5 to presumed normal behavior. Thirty-four raters, on rating 51 patients, agreed 90 per cent of the time even though some of the language used in the description appears to be sufficiently ambiguous and difficult as to require extended training. A similar but briefer device, the Norwich Rating Scales, for

recording patient behavior on disturbed wards, has been developed by Cohen, Malmo, and Thale (2). The ward nurse or attendant checks one of five descriptions of activity, aggressiveness, destructiveness, resistiveness, talkativeness, and tidiness. To check the reliability of the individual scales, 10 patients were rated independently by two raters. The average interjudge coefficient reported was .76.

Rowell (18) has reported a graphic rating scale of 20 items for use by psychiatrically trained nurses. Individual scales are 5-point continua extending from normality on one end to pathology on the other. Variables such as preoccupation, hallucinations, delusions, affect, mood, blocking, and "flight of ideas," are rated. The reported immediate test-retest reliability for 71 total scores is .95, while the correlation between independent ratings by nurses for 62 pairs is .85. The descriptive statements preceding the scale cues appear far too terse and lacking in definition for terms such as blocking and flight of ideas that are ill-defined and notoriously difficult to judge.

The Hospital Adjustment Scale, an ingenious, carefully constructed device for evaluating patient's behavior, has been developed by Ferguson, McReynolds, and Ballachey (11). The scale consists of 91 statements descriptive of psychiatric patients, such as "the patient ignores the activities around him," or "the patient's talk is mostly not sensible." Each statement is marked as True, Not True, or Does Not Apply, for a given patient, and is keyed in such a manner that it is possible to obtain a total score indicative of the patient's general level of hospital adjustment. The scale can be filled out in about 10 minutes by the psychiatric aide or nurse most familiar with the day-to-day behavior of the patient over a period of two weeks to three months. In addition to the total adjustment scores, the scale also offers measures descriptive of (a) communication and interpersonal relations, (b) care of self and social responsibility, and (c) work, recreation, and other activities. The statements can also be grouped as to whether they are indicative of an "expanding" personality, a "contracting" personality, or whether they are neutral. The Hospital Adjustment Scale was developed from a pool of statements descriptive of patients, secured from psychiatric aides. Statements were selected for the final form on the basis of measures of interjudge reliability, ratings made by 16 judges on a scale of over-all hospital adjustment, checks of discriminative power, and percentage of True, Not True, and Don't Know checks. Norms based on the records of 518 patients from four hospitals and clinics are available in percentile form. Patients approaching release from hospital can be differentiated significantly from those judged to be extremely disturbed or chronic hospital residents.

Scherer (20) has prepared a set of 44 four-point scales for the evaluation of patient behavior in such activities as occupational therapy, manual arts, corrective therapy, educational therapy, recreation, and library visits. The Activity Rating Scales were constructed on the basis of a longer list of patient behaviors and tested out on rehabilitation staff personnel. Independent ratings of the same group of patients indicated that, of 1,188 pairs of independent ratings, 60 per cent agree completely, 28 per cent differ by one scale interval, 3 per cent by two scale intervals, and 8 per cent of the scales are not rated. Inasmuch as the activities observed

in rehabilitative situations are similar to noninstitutional, vocational, and social activities, the author postulates that the behavior exhibited may also be similar to and predictive of behavior shown in posthospital adjustment.

SCALES FOR USE BY TRAINED CLINICIANS

The scales and check lists to be described in this section were developed primarily for use by trained clinicians. Most of them represent efforts to secure records of currently observable behavior, symptoms, complaints, or inferable needs and attitudes. A few, however, are intended to secure evaluations of social history obtained from the patient or the patient's friends and relatives.

The Elgin Prognostic Scale, constructed and validated by Wittman and Sternberg (28, 29), is a rating schedule designed to predict recovery in schizophrenics. It consists of 20 rating scales weighted according to prognostic importance; favorable factors are arbitrarily assigned negative weights and unfavorable factors are assigned positive weights. The prognostic score is the algebraic sum of the weighted measures. Most of the variables, such as shut-in personality, type of onset, or range of interests, are based upon premorbid social history secured from the patient's relatives or from the patient himself. A few are based on currently discernible symptomatology such as hebephrenic symptoms, ideas of influence, bizarre delusions, or affect. These scales were constructed following a review of available literature on prognostic factors in schizophrenia. They were validated and cross-validated on Elgin State Hospital patients, and shown to predict outcome of shock treatment with greater accuracy than staff judgments. A multiple-factor analysis (6) of intercorrelations between the scales based on a group of 200 patients revealed three interpretable factors: schizoid withdrawal, schizophrenic reality distortion, and personality rigidity or inadaptability. The principal shortcoming of this device is the difficulty of securing reliable premorbid personality pictures from parents and relatives.

Saslow and his associates (19), in their effort to study the personality correlates of psychosomatic disorders, have prepared some 12 scales for measuring habitual patterns of reactions to crises. Each pattern is briefly delineated by means of a 5-point scale. Ratings are based on social history data secured from patients during psychiatric interview. Included are scales of impulsiveness, subnormal assertiveness, obsessive-compulsive behavior, depressive behavior, anxiety, hysteria, inward expression of emotions, low awareness of body symptoms, insecure feelings of inferiority, repressed hostility, and strong dependent needs. No reliability data are reported for the individual scales.

A check list resembling the Phipps Clinic Chart, but more systematic in its approach, has been developed by Peters (16). The check list, which consists of 199 traits, is grouped under 7 categories: history, acting, talking, mood, emotion, interests, ideation. The traits were compiled from interview records and presumably cover a large proportion of all characteristics required to describe a patient's personality and symptoms. Three ratings are used: a plus for a positive degree of a trait, no mark at all when a trait is present to a normal degree or does not pertain to the subject, and a minus sign for a negative degree of the trait. Ward behavior, interview data, and social history

may be taken into account in rating. The check list has been used successfully to identify traits related to improved adjustment of lobotomized patients (16).

A mental health check list entitled The Pattern for Living has been constructed and reported by Conrad (3). The check list is designed to measure positive mental health, social conformity, and pathology. It is for use by trained personnel in appraising persons applying for and receiving outpatient psychotherapy. Every item regarded as true is checked plus, false items are checked minus, while the remainder may be scored by a question mark. Trends are indicated in a separate column. Of the 45 items, 16 are concerned with positive mental health, 12 with conformity, and 17 with pathology. Evidence is presented that patients with high positive mental health scores tend to stay in therapy (3).

A Guided Clinical Interview Analysis for use in connection with structured clinical interviews has been reported by Abt (1). Eight scales descriptive of attitude toward parents, attitude toward siblings, attitude toward childhood, attitude toward people in general, and attitude toward sex make up the Analysis. Each subcategory of a scale is ratable on a 5-point scale. Agreement between raters for recorded interview material is reported to be high. A scale similar to the Abt form as to content, for use in evaluating adult outpatients receiving psychotherapy, is being developed by Morse (15). The six major areas measured are accessibility to therapy, occupational and school adjustment, social adjustment, sexual adjustment, family adjustment, and symptomatology. Most of the 40 items contain four brief descriptions to characterize grades of behavior on the scale con-

tinuum. Thus far neither reliability nor validity data are available.

The Psychiatric Rating Scales developed by Malamud *et al.* (13), at the Worcester Hospital in Massachusetts, represent yet another rating form for the quantitative recording of psychiatric clinical findings and the changes that occur in them during the course of illness. The scale consists of 19 "functions" divided into three major groups. The first seven functions comprise behavior items, such as sexuality and sleep, which depend upon continuous observations by ward personnel during the 24 hours preceding the psychiatric ratings. The third group consists of eight functions that are evaluated during the interview and require verbal reactions from the patients themselves. Associations, memory, and thought processes are examples of these eight functions.

Each of the 19 scales extends to the left and to the right from a central base line consisting of two terms which represent the usual variations in the particular function that is within normal limits. On both sides of the base line are indicated the range of deviation in terms of progressive degrees of pathology. On the left side of the scale are those deviations which are directed toward the outside (centrifugal). To the right of the base line are those centripetal or internally directed deviations in the function. Steps in the function are identified by three psychiatric terms to the left of the base line and three to the right. Ratings made for either half of a function may range from 1 to 6. The Feeling function, for example, is marked by the following sequence of terms: panic, anxiety-guilt, tense-irritable, hypersensitive, hyposensitive, phlegmatic, dull, apathetic. A patient's total score consists of the sum of his scores on each

of the 19 functions. A correlation of .92 was secured between 100 paired independent ratings made by two psychiatrists on 26 patients. Application of the Psychiatric Rating Scale to agitated depressives (12) indicates a good correlation with changes in the clinical picture resulting from electroshock therapy. The scale has also been used to record changes resulting from prefrontal lobotomy.

A criticism that may fairly be directed at the Psychiatric Rating Scales is that the terms used to describe the 19 functions are nowhere defined. This would suggest that ratings on individual functions are less reliable than the total scores. In total scores, differences between raters are cancelled out in the process of summing. Lockwood (5) reports that the evaluations of his psychiatric raters as reflected on individual scales were not sufficiently reliable for use as criteria for clinical improvement. In a number of the scales several variables seem to have been forced into a single bipolar continuum. Thus, instead of one rating there should be two on such functions as Sexuality. It would have been preferable to determine empirically in advance whether or not presumably antithetical forms of behavior were mutually exclusive.

The Elgin Behavior Rating Scale Revised has been developed by Wittman and Hills (29) for the purpose of describing psychiatric patients in six areas of behavior. A rating is accomplished by selecting the descriptive sentence listed under a category which is most applicable to the patient. The scales are graded from "very poor" to "very good." A weight of 0 is given to the most deviant behavior; intermediate steps are assigned weights from 1 to 4. Somatic behavior is covered by seven scales descriptive of physical appearance,

physical condition, appetite, and the like. The Social Behavior area consists of six scales descriptive of such aspects as conversation, cooperation, and sex behavior. There are eight scales to describe Mental Behavior such as orientation, insight, and affective response. Under the rubric of Psychotic and Neurotic Behavior are three global scales descriptive of affective exaggeration, paranoid projection, and schizoid withdrawal. Neurotic Behavior and Anti-Social Behavior are separately rated on two scales. No data with regard to inter-rater agreement are reported. Most of the scales contain several variables for rating. The orientation scale, for example, includes items relative to disorientation as to time, place, and person. There are actually 63 separate variables available for rating. Although no data are provided by the authors, the total score would probably represent a useful measure of over-all severity of illness.

A series of check lists designed to describe patient character, temperament, and intellectual capacity have also been developed by Wittman (29) and her associates. The check list of Fundamental Temperament Reactions postulate three bipolar components. The first component, affective exaggeration, extends from manic expansion to depressive constriction. Aggressive ascendance and defensive passivity mark the two ends of a continuum of paranoid compensation. The schizoid withdrawal component extends from heboid regression to simple withdrawal. Two descriptive elements identify manic expansion and a similar number are used to describe depressive constriction. The rater may check as few or as many elements as he observes; the total number of checks provides a total score. Wittman has also constructed check lists for Temperament

Deficiencies, Anxiety Reactions, Character Deficiencies, Exaggerated Reaction Types, Addictive Reaction Types, Disorders of Intellectual Capacity, Constitutional Intelligence Disorders (Amentia), and Acquired Intelligence Loss (Dementia).

Wittenborn (23) has reported a rating-scale procedure for the evaluation of mental hospital patients as one step in a broad program directed at the development of a quantified method for multiple psychiatric diagnoses. The procedure was devised to permit a psychologist, psychiatrist, nurse, or other competent observer to (a) rate the currently discernible symptoms of a psychiatric hospital patient, (b) score these ratings, and (c) prepare a profile which would indicate to what extent the patient's pattern of symptoms resembled the symptom patterns found among psychiatric hospital patients generally. The rating schedule consists of 55 different, unlabeled scales presented sequentially in a random manner. The scales were selected to represent a fairly adequate sample of important symptoms that characterize hospitalized patients. They were designed to demand a minimum of interpretation and experience from the observer, and to yield judgments which are relatively unbiased by the rater's particular theory or point of view. The directions require that the rater check the most pathological condition or level of behavior observed during the period studied and that every scale be checked for every patient. This last procedure undoubtedly facilitates correlational studies. However, there is a real question as to how much error is introduced if judgments are forced as in the case of the mute patient or the patient who is evasive or irrelevant in his speech. A mute patient, for example, must be judged as to rate of change of ideas, insight, or rate of speech.

Most of the scales consist of four elements; a lesser number contain three or five elements. The rating process, which consists of encircling one element for every scale, requires about 10 or 15 minutes.

In an effort to determine the existence of clusters on patterns of symptoms and the influence of such characteristics of the sample as age, sex, or organic brain damage on these patterns, Wittenborn (24, 25) has completed a series of factorial analyses of his rating schedule. The nine factors or psychiatric syndromes which have been repeatedly identified are: Acute Anxiety, Conversion Hysteria, Manic State, Depressed State, Schizophrenic Excitement, Paranoid Condition, Paranoid Schizophrenic, Hebephrenic Schizophrenic, and Phobic-Compulsive. Scoring weights for the symptom scales and norms for transmuting cluster scores into standard cluster scores are available for the Descriptive Scales for Rating Currently Discernible Psychopathology.

The Multidimensional Scale for Rating Psychiatric Patients (MSRPP) consists of two sets of brief, relatively objective rating scales for the description of the behavior, symptoms, complaints, and inferred motivation of psychiatric patients. One form or set is for outpatient use and the other for hospital use. These schedules represent a systematic effort to (a) develop a quantified record or description of mentally ill patients that could be used to measure change or improvement, and (b) isolate and identify underlying unitary variables. The form for outpatient use, developed by Lorr, Jenkins, Holsopple, and Rubinstein (8), consists of 49 unlabeled, randomly presented, 4- and 6-point graphic rating scales for describing outpatients as seen in diagnostic or therapeutic interview. It is intended only for use by psychologists or psy-

chiatrists. The scales, calling for judgments based on directly observable manifest behavior, are grouped together, as are those which require inferences. Included are scales descriptive of personality traits such as emotional responsiveness, complaints such as headaches, and symptoms such as compulsive behavior. The schedule provides scores for 16 factors identified on an original set of 73 (9). Tentative norms and standard scores for profiling individual patient records are available. The factors measured by this form have been labeled Hostility, Reality Distortion, Obsessive-Compulsive Reaction, Sex Conflict, Gastro-Intestinal Reaction, Cardiorespiratory Reaction, Anxiety-Tension, Anxious Depression, Emotional Responsiveness, Adaptability, Sense of Personal Adequacy, Vigorous Interest, Conscientiousness, Independent Maturity, Goal-Directed Motivation, and Prudence.

The MSRPP form for hospital use represents a revision of the Northport Record, which was initially constructed and developed by Lorr, Singer, and Zobel (7). It consists of 50 unlabeled graphic rating scales presented in a random order. The schedule was designed to measure the major symptoms characteristic of recognized syndromes of the various psychoses and behavior readily observable in a routine diagnostic interview or on the ward by nurses and psychiatric aides. A rating is made by encircling the entry which is most typical or representative of the patient during the period observed. When there is no basis for rating a patient on a particular scale, an unratable category is encircled.

The hospital form of the MSRPP provides tentative norms, based on 450 patients from four psychiatric hospitals, for 12 factors or syndromes measured. These factors were identified in a multiple-factor analysis of 55 of the original 81-item Northport Record. Standard score measures are available for profiling a patient on the following factors: Manic Excitement, Retarded Depression, Anxious Depression, Perceptual Distortion, Conceptual and Thinking Disorganization, Paranoid Suspicion, Grandiose Expansiveness, Schizophrenic Excitement, Disorientation, Withdrawal, Hostile Aggressiveness, and Activity Level. The last three of these factors are behavior parameters observed on the ward.

The sum of the absolute deviations from a "normal" pattern provides an over-all index of severity of illness which has been found to reflect change resulting from lobotomy. Seven type patterns similar to conventional diagnostic description are also available for use as an aid in diagnosis.

DISCUSSION

The development and use of rating scales and check lists for the systematic recording of clinical judgments of manifest behavior and inferred attitudes and needs appear to represent an important advance for the clinician and the research worker.

There seems to be no doubt that the interview is here to stay even though some critics, particularly the psychometricians, would have it replaced by other and presumably more rigorous procedures. The problem becomes one of developing controlled interview patterns as suggested by Zubin (31) and of objectively recording what the trained clinician can validly or reliably observe or infer. While the new techniques for sound recording of interviews are unquestionably important for evaluation, they are not substitutes for the clinician nor do they provide any complete basis for the analysis of the interview. Visual as well as auditory cues provide important data for an

appraisal. We are simply saying, in brief, that the clinician can contribute toward the description of his patient and the prediction of his future behavior. The rating form provides a framework for quantifying his judgments, for jogging his memory, and for minimizing "halo" bias.

The rating schedule also offers a common conceptual framework for the clinician regardless of the examination procedure used. Clinical judgments derived from an analysis of the Rorschach test, the TAT, or a sentence completion form may be recorded in objectified form on the rating scales. Ratings can be useful in defining and clarifying areas of agreement and disagreement. Clinicians differing in theoretical orientation can find a common ground when a concept characteristic of an individual is stated simply, in graded form. When defined in simple understandable terms, many presently elusive and amorphous variables can be checked for reliability and related to a larger domain of objectively expressed concepts. Conceptual formulations often loosely used, such as sexual identification or ego strength, can be pinned down for closer scrutiny and validation.

In any field where agreement on basic variables is lacking, factor analysis is a powerful tool for resolution of complex concepts into simpler elements and for the identification of underlying parameters. The confusion in, and duplication of, vocabularies for describing personality and psychodynamics are especially notable. However, Moore, Wittenborn, and others (14, 25) have shown that rating scales and check lists can be utilized fruitfully for the isolation and identification of psychopathological syndromes and categories. Investigators concerned with the isolation of primary factors in perception and cognition have recently recommended a battery of tests to represent each of the better defined factors in future factorial studies. In the absence of more objective measures of personality and psychopathology, it may be useful to utilize similarly, as reference variables, standard sets of rating scales in further factorial investigations.

SUMMARY

The purpose of this review has been to examine and report on rating scales and check lists designed to describe psychiatric patients in the interview and on the ward that have appeared during the past ten years. A half dozen scales and check lists suitable for use by nurses and psychiatric aides is reported in the literature. Of these devices the Hospital Adjustment Scale has been most carefully developed and seems to be usable by any psychiatric aide who can read. For more precise measurement of change the Gardner Behavior Chart and the Northampton Activity Rating Scale may be preferable.

Among those designed for use by psychologists and psychiatrists to describe psychotic behavior and symptomology, the Wittenborn Descriptive Scales and the Multidimensional Scales have been most intensively analyzed and developed. The Elgin Prognostic Scales are at present the most useful for predicting improvement although difficult to use because of the practical problem of securing reliable data on the patient's past history.

The rating schedule offers considerable promise as a procedure for quantifying the interview, for isolating basic psychopathological parameters, and for generally providing a conceptual basis against which clinicians of varying persuasions and training can find a common ground.

REFERENCES

1. ABT, L. E. The analysis of structured clinical interviews. *J. clin. Psychol.*, 1949, 5, 364.

2. COHEN, L. H., MALMO, R. B., & THALE, T. Measurement of chronic psychotic overactivity by the Norwich rating scale. *J. gen. Psychol.*, 1944, 30, 65–74.

3. CONRAD, D. C. Towards a more productive concept of mental health. *Ment. Hyg.*, 1952, 36, 456–473.

4. KEMPF, E. J. The behavior chart in mental diseases. *Amer. J. Insanity*, 1915, 71, 761–772.

5. LOCKWOOD, W. L. Some relations between response to frustration (punishment) and outcome of electric convulsive therapy. *Comp. Psychol. Monogr.*, 1950, 20 (Ser. No. 104), 121–186.

6. LORR, M., WITTMAN, PHYLLIS, & SCHANBERGER, W. An analysis of the Elgin prognostic scale. *J. clin. Psychol.*, 1951, 7, 260–262.

7. LORR, M., SINGER, M., & ZOBEL, H. Development of a record for the description of psychiatric patients. *Psychol. Serv. Cent. J.*, 1951, 3, No. 3.

8. LORR, M., RUBINSTEIN, E., & JENKINS, R. L. A factor analysis of personality ratings of outpatients in psychotherapy. *J. abnorm. soc. Psychol.*, 1953, 48, 511–514.

9. LORR, M., SCHAEFER, E., RUBINSTEIN, E. A., & JENKINS, R. L. An analysis of an outpatient rating scale. *J. clin. Psychol.*, 1953, 9, 296–299.

10. LUCERO, R. J., & MEYER, B. T. A behavior rating scale suitable for use in mental hospitals. *J. clin. Psychol.*, 1951, 7, 250–254.

11. MCREYNOLDS, P., BALLACHEY, E. L., & FERGUSON, J. T. Development and evaluation of a behavioral scale for appraising the adjustment of hospitalized patients. *Amer. Psychologist*, 1952, 7, 340. (Abstract)

12. MALAMUD, W., HOAGLAND, E., & KAUFMAN, I. C. A new psychiatric rating scale. *Psychosom. Med.*, 1946, 8, 243–245.

13. MALAMUD, W., & SANDS, S. L. A revision of the psychiatric rating scale. *Amer. J. Psychiat.*, 1947, 104, 231–237.

14. MOORE, T. V. The essential psychoses and their fundamental syndromes. *Studies in psychology and psychiatry.* III. Baltimore: Williams & Wilkins, 1933.

15. MORSE, P. W. Proposed technique for the evaluation of psychotherapy. *Amer. J. Orthopsychiat.*, in press.

16. PETERS, H. N. Traits related to improved adjustment of psychotics after lobotomy. *J. abnorm. soc. Psychol.*, 1947, 42, 383–392.

17. PLANT, J. S. Rating scheme for conduct. *Amer. J. Psychiat.*, 1922, 1, 547–572.

18. ROWELL, J. T. An objective method of evaluating mental status. *J. clin. Psychol.*, 1951, 7, 255–259.

19. SASLOW, G., GRESSEL, G. C., SHOBE, F. O., DUBOIS, P. H., & SHROEDER, H. A. Possible etiologic relevance of personality factors in arterial hypertension. *Psychosom. Med.*, 1950, 12, 293–302.

20. SCHERER, I. W. A behavior rating scale for use in activity therapy situations. *Info. Bull.*, Dep. Med. & Surg., Psychiat. & Neurol. Div., Veterans Administration, Jan., 1951.

21. SCHRADER, P. J., & ROBINSON, M. F. An evaluation of prefrontal lobotomy through ward behavior. *J. abnorm. soc. Psychol.*, 1945, 40, 61–69.

22. WILCOX, P. H. The Gardner behavior chart. *Amer. J. Psychiat.*, 1942, 98, 874–880.

23. WITTENBORN, J. R. A new procedure for evaluating mental hospital patients. *J. consult. Psychol.*, 1950, 14, 500–501.

24. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, 15, 290–302.

25. WITTENBORN, J. R., & HOLZBERG, J. D. The generality of psychiatric syndromes. *J. consult. Psychol.*, 1951, 15, 372–380.

26. WITTENBORN, J. R., MANDLER, G., & WATERHOUSE, I. K. Symptom patterns in youthful mental hospital patients. *J. clin. Psychol.*, 1951, 7, 323–327.

27. WITTENBORN, J. R., BELL, E. G., & LESSER, G. S. Symptom patterns among organic patients of advanced age. *J. clin. Psychol.*, 1951, 7, 328–331.

28. WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, 4, 20–33.

29. WITTMAN, PHYLLIS. The Elgin check list of fundamental psychotic behavior reactions. *Amer. Psychologist*, 1948, 3, 280. (Abstract)

30. WITTMAN, PHYLLIS, & STERNBERG. L. Follow-up of an objective evaluation of prognosis in dementia praecox and manic-depressive psychoses. *Elgin Pap.*, 1944, 5, 216–227.

31. ZUBIN, J. Objective evaluation of personality tests. *Amer. J. Psychiat.*, 1951, 107, 569–576.

RECENT STUDIES OF SIMPLE REACTION TIME[1,2]

WARREN H. TEICHNER

Aero-Medical Laboratory, Wright-Patterson Air Force Base

In spite of the important role that the human reaction time study has played in the development of psychological science and the tremendous amount of research effort expended in its behalf, there are still large gaps in our knowledge of the empirical relationships in which reaction time is involved. This report is an assessment of the present scientific status of the topic based primarily on the experimental literature of the last twenty years. Previous reviews were presented by Woodworth (161) in 1938 and by Johnson (82) in 1923. In addition, Forlano, Barmack, and Coakley (50) have reviewed the effects of ambient and body temperature on both simple and choice reaction time and Finan, Finan, and Hartson (46) have briefly summarized the use of reaction time scores as measures of performance decrement.

First, it should be noted that there are several kinds of reaction time experiments (8, 161). Because of the

tremendous literature involved, the present discussion will be restricted to the *simple reaction time* (RT), which is the time interval between the onset of the stimulus and the initiation of the response under the condition that S has been instructed to respond as rapidly as possible. In order of presentation this review will consider the effects on the RT of stimulus and receptor conditions, central and motor factors, and certain special conditions such as the effects of low ambient temperature, loss of sleep, etc.

Next, the complexity of the measure should be recognized. After the onset of the stimulus there is a lag, or latent period, during which the receptor process is initiated and builds up to a maximum (25, 56, 119, 120, 140, 141). This is followed by a second lag involving central transmission of the sensory impulses to the motor fibers, and finally, there is a time delay involved in the contraction of the muscles (33, 34, 47, 67, 127) and the beginning of the movement of the responding member. Any of the factors which affect any of these processes will obviously also affect the measured RT within which they are present. Since most RT studies deal only with the over-all measure of time, the present discussion will be confined to this measure. However, a discussion of sensory latent periods relevant to the visual RT may be found in Arnold and Tinker (2) and in Strughold (140, 141), to the auditory RT in Chochell (24, 25), to the pain RT in Pattle and

Weddell (117), and to the thermal RT in Wright (162). Piéron (120) summarizes the topic in considerable detail.

STIMULUS-RECEPTOR FACTORS

RT as a Function of the Sense Modality Stimulated

It is a common assumption based on the neuro-anatomical differences existing among the various receptor systems that the time of reaction varies according to the sense modality stimulated. Textbooks usually contain comparisons between the RT's obtained by various kinds of stimulation, frequently presenting lists in which RT is ranked according to the senses. However, little attention appears to have been given to the logic of measurement involved in making such comparisons. For example, in order to say that the auditory RT is shorter than the visual one, as is usually done, the two types of stimulation must be compared on the same scale. The scales that have been employed are, unfortunately, scales of subjective intensity and these cannot be considered comparable from one sense modality to another. This argument, i.e., that the attribute terms of one psychophysical dimension logically cannot be projected to another such dimension, represents what is usually thought of as an advance in the logical foundation of psychology and will be remembered as having invoked considerable discussion.[3] Although the problem was resolved with regard to the comparing of sensations, it seems to have been ignored in dealing with the RT. For this reason, the conclusion that must be drawn is that *there is no evidence available that indicates whether or not the RT varies according to the receptor system stimulated.*

The experimental literature available does allow one kind of meaningful conclusion regarding this matter. Where studies have been done comparing a specific intensity of, say, sound with a specific intensity of, say, light, it should be possible to decide whether the RT's for those specific values, and those only, are shorter for the one sense than for the other. Unfortunately, most studies making such comparisons either report no intensity values at all, or they report them in arbitrary units with no means of reference.

It is possible, of course, to speculate. The literature concerned with visual and auditory RT's is almost universal in reporting faster RT's to the sound stimuli than to other stimuli. Most studies have also found faster RT's to tactual than to visual stimuli. Robinson (125), for example, presented a summary table of eight of the older investigations in which the RT's for vision, hearing, and touch were compared. If medians are calculated from this table, the RT for audition is found to be 0.142 sec., for touch 0.155 sec., and for vision 0.194 sec. In all eight of the experiments, the auditory RT was consistent in being faster than the visual one. But in four of the eight, the tactual RT was faster than the auditory one and in the other four the opposite result was obtained. Todd (146), and in addition more recent studies (4, 16, 38, 48, 153, 154), concur in finding shorter RT's to sound than to light. One other study (Moldenhauer) reviewed by Woodworth (161) found the auditory RT faster than the tactual one. Lanier (92), however, in a study of the effect of training on auditory, visual, and tactual RT's found the

---

[3] For a complete statement of this problem see Boring (10).

tactual RT shortest for trained Ss with the auditory and visual RT's being approximately equal. With untrained Ss, on the other hand, the auditory RT was shortest with visual and tactual times being equal.

It is possible that the speed of reaction with respect to sensory modality may depend on some sort of speed or reaction factor which determines the kind of stimulus to which an individual will exhibit the shortest RT. Wells, Kelley, and Murphy (153, 154) studied the ratio of the median RT to light to the median RT to sound and found a light-sound ratio of 1.15 for 11 untrained Ss. Two trained Ss exhibited ratios of 1.34 and 1.45. They also found a correlation of −.52 between the ratios and the median RT to sound in the untrained group. From this they concluded that Ss who have a relatively fast RT to sound have a relatively slow RT to light, and vice versa. The ratios obtained agree fairly well with those recently reported by Canfield, Comrey, and Wilson (16). However, the correlation of −.52 has little meaning since it is logically possible to obtain a negative correlation between a ratio and its denominator even when the direct correlation between the numerator and the denominator is positive. In fact a negative correlation merely indicates that the ratio is greater than 1.00. More direct comparisons by Forbes (48) and by Lanier (92) have revealed positive correlations of .48 and .90 respectively between response to light and to sound.

From a consideration of chemical, temperature, pressure, or electrical stimuli applied directly to the skin to elicit pain response, Woodworth (161) concluded that the slowest RT is that based on painful stimulation. Recent studies have also recorded pain RT's to radiant heat (e.g., 61, 139, 156)

and radiant cold (35). It is necessary, however, to consider not only the measurement problems involved but also the logical and technical difficulties involved in the measurement of pain (37, 61).

Wright (162) investigated the RT's associated with cutaneous sensations of warmth elicited by visible and infrared radiation. In 22 out of 39 men tested, RT was faster to stimulation on the back of the hand than it was on the palm. In 17 out of 27 others who were tested, stimulation in the epigastrium produced faster RT's than stimulation in the interscapular region. In 78 Ss the RT to light was found to vary from about 0.33 sec. for a very intense sensation to about 20 sec. at threshold. One interesting result was that psychophysical functions obtained with RT measures showed a qualitative similarity to visual functions of intensity, duration, and area.

A little work has also been done recently on proprioceptive RT's. Chernikoff and Taylor (21) measured the speed of reaction to the sudden falling of the S's arm. When the response measure was the time of release of a telegraph key, no differences were obtained between auditory, tactual, and kinesthetic RT's. However, when the response was the stopping of the arm movement, RT was considerably shorter than the key-release type of reaction regardless of whether the latter was to an auditory, tactual, or kinesthetic stimulus. Hick (72) and Vince (152) have also studied the RT's involved in making corrective movements in a pursuit task involving both visual and proprioceptive components.

In a different type of proprioceptive study, Baxter and Travis (5) measured the vestibular RT's of 31 Ss to rotary motion of the body. The Ss were blindfolded, seated in a con-

stant-speed revolving chair, and instructed to press a telegraph key upon detecting a change in motion. When the chair was moved from a stationary position, a mean RT of 0.52 sec. was obtained; when the chair was in motion and the direction of motion was changed, a mean RT of 0.72 sec. was obtained. From this highly significant difference Baxter and Travis concluded that the RT to perception of motion from rest is faster than the RT to perception of change in direction of motion.

The possible variation of RT with the sense modality used for stimulation is a question which has not been answered by the studies reviewed. As noted above, genuine comparisons require the use of common scales and such scales have not been employed. One possibility presently available is to determine the RT's on a statistical or probability basis and make comparisons in terms of a probability of RT scale. The first step is to determine empirically for each sensory modality the relationship between RT and the intensity dimension used for that modality, i.e., the function $P(RT) = f(I)$. This relationship is presumably sigmoid. Time of reaction (RT) can now be determined as a function of the probability of reaction for each case, i.e., the function, $RT = g[P(RT)]$ can now be obtained. Since $P(RT)$ is common for all modalities, this function furnishes a legitimate basis of comparison within the range, $0 < P < 1.00$.

RT as a Function of the Number of Sense Organs Stimulated

No recent work has been done on mono- versus bisensory stimulation. As indicated in previous reviews Poffenberger (121) reported the RT to light about 0.015 sec. faster for each of the three Ss he used when both eyes were stimulated than when only one was stimulated. This is hardly conclusive considering the size of his sample and the likely error of measurement. Similarly, Bliss (9) found the RT slightly shorter for binaural than for monaural stimulation. More recently, Smith (136) found faster RT's to the apparent movement of objects when viewed binocularly than when viewed monocularly.

Since both visual and auditory phenomena usually show differences according to whether the stimulation is mono- or bisensory, it is reasonable that RT's should show corresponding differences. The high expectation of these differences is probably the reason why experimenters have been so little motivated to provide demonstrations of them. This is unfortunate because neither Poffenberger's data nor those of Bliss should be considered reliable enough to be sufficient.

Simultaneous presentation of stimuli to different senses has also been studied. Todd (146) presented light, sound, and electric shock singly and in simultaneous combinations and measured the speed of simple reaction to each. Combined stimuli in every case elicited faster RT's than the individual stimuli making up the combination. The RT to combined sound and light, for example, was not only faster than to light alone but also faster than to sound alone. The shortest RT's were to the combination of all three types of stimulus. On the other hand, *successive* stimulation of different sense organs produced longer RT's than did stimulation of single sense organs.

RT as a Function of Number of Receptors Stimulated

According to neurological principles of summation, it would be expected that the greater the number of receptors stimulated, the shorter the

latent period and, consequently, the shorter the time of reaction. There are two studies which support this hypothesis. Older data from Froeberg (55) provide some support in the visual case. These data show that, as the retinal area stimulated is increased from three to 48 sq. mm., RT decreases from 0.195 sec. to 0.180 sec., the function appearing to be negatively accelerated and still decreasing at 48 sq. mm. Wright (162) reported a similar type of phenomenon for thermal RT's.

RT as a Function of the Location of the Stimulation in the Visual Field

The visual RT varies with the portion of the visual field stimulated, according to Poffenberger (121). This investigator stimulated at 3°, 10°, 30°, and 45° away from the fovea and measured the increase in length of RT over the RT obtained from foveal stimulation. He found that RT increased in the temporal periphery from about 0.004 sec. at 3° to about 0.024 sec. at 45°, and in the nasal periphery from about 0.004 sec. to about 0.015 sec. Except for stimulation at 3°, increases in RT were consistently greater for temporal as compared to nasal stimulation.

Poffenberger's data suggest that RT is positively correlated with ability to perceive shape since shape perception is greatest in the foveal area and decreases toward the periphery. RT seems to be correlated also with visual acuity (reading of numbers, letters, etc.) since this too decreases with distance away from the fovea and is greater in the nasal half. These relationships, although not recently confirmed, present some very interesting possibilities for the field of visual measurement and suggest, among other things, speed-of-seeing techniques for the testing of visual acuity. Such techniques have had

some use in the measurement of acuity as related to the amount of illumination (26, 45, 96).

Longer RT's with stimulation in the peripheral area would not be expected in the dark-adapted eye, peripheral sensitivity to weak light being greater under this condition. Data relevant to this last hypothesis were obtained by Lemmon and Geisinger (94) who measured the RT's of 14 Ss at the fovea and 45° away. Under light-adapted conditions they found the RT significantly longer in the periphery, which supports Poffenberger's (121) results. When the eye was dark-adapted, they found the average RT of their Ss slightly shorter in the periphery, which is in accord with the hypothesis. This result, however, was not statistically significant.

RT as a Function of the Intensity of the Stimulus

Considerable research and theoretical effort have been expended on the relationship between RT and stimulus intensity. Early studies (7, 20, 26, 45, 55, 56) all agree that the visual RT becomes shorter as the intensity of light is increased. More recent investigations (32, 96, 137, 138) concur. In spite of earlier controversies, there is little doubt that the relationship is a nonlinear one, not only in the case of the visual RT but also for auditory (24, 25, 118), gustatory (118), thermal (162), and pain (35, 73, 120) RT's. Attempts have been made to fit the intensity data into mathematical, theoretical frameworks, with exponential, hyperbolic, and parabolic functions all being used more or less successfully on the same sets of data (78, 91, 118, 124).

In all the intensity studies cited, the intensity of the stimulus was varied and the speed of reaction measured when the receptor was in a "normal"

condition. In a related, but somewhat different type of study, Hovland (76) carefully investigated the effect on the RT to a test stimulus of 250 foot-candles of having the eye previously adapted to lights ranging from zero to 200 foot-candles. As might be expected, RT became shorter as the difference between adaptation brightness and stimulus brightness increased. This result is not altogether clear, however, since Lemmon and Geisinger (94), who also used a very bright test stimulus, found shorter RT's with the light-adapted than with the dark-adapted eye.

In addition to varying the intensity of the stimulus or of the adapting stimulus, it is also possible to vary the magnitude of change in the intensity of on-going stimulation. The RT has been studied in this way only in vision. Steinman (137) found that the RT became shorter as the relative magnitude of the change in intensity was increased. The relationship held only up to a limit, however, after which the RT began to increase again. This suggests that the function is not monotonic, but rather has an optimum at some moderate amount of change in intensity. If this conclusion is supported, it may provide an explanation for the discrepancy between Hovland's (76) finding a consistent decrease of visual RT with greater differences between adaptation intensity and test stimulus intensity and the opposing result obtained by Lemmon and Geisinger (94); i.e., the change in stimulus intensity in the latter study may have been at or beyond that point in the function where RT begins to become longer. Some support for this hypothesis is to be found in a study reported by Johnson (81). In this experiment Johnson darkened one half of a photometric field by about 4.6 per cent and compared the RT's

when the surround was 2.25 times as bright as the test field, 0.75 times as bright as the test field, or too dark to be measured. The test field itself had a brightness of 7.8 millilamberts and was of foveal dimensions. Johnson reports that the slowest RT was that elicited under the darkest condition, the next slowest under the brightest condition, and the fastest RT was obtained under the moderately bright surround. All differences were statistically significant. These results, along with those of Steinman, indicate the likelihood of an optimum intensity change and again suggest an explanation for the otherwise contradictory results of Lemmon and Geisinger.

RT's have been used effectively to study the effect of illumination on visual acuity. Luckiesh (96, p. 131) and Cobb (26) both found that speed of vision (RT) increases rapidly with increases in illumination up to about 18 to 20 foot-lamberts, after which further increases in illumination have no significant effect. Data from Ferree and Rand (45) indicate that speed of seeing is also a function of size of test object. In this study the limit of speed reached a maximum at approximately 25 foot-candles for relatively large test objects (3, 4.2, and 5.2 minutes of visual angle) and at approximately 45 foot-candles for relatively small test objects (1 and 2 minutes of visual angle).

RT as an Index or Measure of Sensation

The study of the effect of changes in magnitude of stimulus intensity on RT suggests the possibility of applying RT measures to psychophysical problems. The use of choice reaction times in what has been called the method of judgment time (17, 40, 144) is not new, of course. Although the use of the RT as a psychophysical

measure has always been implicit in the design of experiments involving visual and auditory thresholds, it has only recently had serious employment in this regard.

Steinman (137) studied the adequacy of the RT to change in brightness as a psychophysical method. As discussed above, he found that RT decreased as a (seemingly hyperbolic) function of the magnitude of change. With a constant stimulus-ratio this relationship was maintained up to the higher intensity levels where a reversal occurred. Although he attributed this reversal to an adaptation effect, other hypotheses are possible, as was indicated in the discussion of the effect of intensity.

Steinman (137, 138) also observed that the RT was faster to a decremental change in magnitude than it was for an objectively equal increment of change. This places a restriction on the method, but it is a restriction not without parallel in the standard psychophysical techniques. In any case, Steinman was able to plot RT functions which were in close agreement with similar functions obtained by more customary procedures, and for this reason he concluded that the method is adequate for securing equal perceptibility contours, threshold measurements, etc.

Other similar studies have been done, most of them quite recently. Galifret and Piéron (57) were also successful in obtaining visual functions based on differences among RT's. Chocholle (25) discusses the problem relevant to the psychophysiology of hearing; Wertheimer (156) suggests using the RT for obtaining radiant heat pain thresholds. Essentially this has been done both for the determination of the radiant heat pain threshold (61, 139) and the radiant cold pain threshold (35). Wright (162), furthermore, has obtained the

Weber type of function by using cutaneous RT's as an index of sensations of warmth produced by light radiation.

RT as a Function of the Duration of the Stimulus

It is difficult to see why the duration of the stimulus should influence the RT to the *onset* of a suprathreshold stimulus unless some type of summation of intensity hypothesis is advanced. Nevertheless, there is some suggestion in the literature that stimulus duration does have an effect.

Froeberg (55) varied visual stimuli by equal geometric intervals between 0.003 sec. and 0.048 sec. Within this range he found that the longest durations produced the shortest RT's, the function (of the geometric intervals being linear.

Wells (155) varied the duration both of sound and of light stimuli. For an auditory stimulus he used the sound of an electric buzzer with durations of 0.007, 0.036, 0.051, 0.076, and 0.106 sec. Two *S*s gave 200 responses under each duration. The results indicated that RT was a linear function of the logarithm of the duration of the stimulus, which is in agreement with Froeberg's visual data. But, although the form of the relationship was the same as that obtained by Froeberg, *the slope was in the opposite direction*. To study the effect of duration in the visual case, Wells used a constant intensity stimulus (brightness of 0.12 millilamberts) at five durations ranging between 0.012 and 1.00 sec. In this experiment *S*s responded to the onset of the light. In a second experiment performed with the same *S*s, response was made to the cessation of the light. Five durations of light were used again, this time ranging between 0.010 and 1.00 sec. The results differed from those of Froeberg in that they indi-

cated that there is an optimal duration and that this duration varies from individual to individual. Whatever the individual optimum, RT tended to become longer with deviations from it. The range of optima for ten *S*s was between 0.025 and 0.066 sec. However, large variations were found not only for the optima among *S*s, but in the optima between stimulus onset and cessation even for individual *S*s; i.e., the optimum was usually not nearly the same for a single *S* under the two conditions.

Brogden and his students (21, 22, 62, 63) recently performed a series of experiments in which they compared the RT to auditory stimuli of fixed duration with the RT for response-terminated durations. They concluded that the primary variable is stimulus duration and that response termination merely acts on this factor. However, they also found that for longer sound durations (400–2000 msec.) the RT to the response-terminated stimulus was slightly, but significantly, shorter than to the fixed duration stimulus. For the shorter durations (100–200 msec.) there were no differences. Scrutiny of their data leaves the impression that RT increased in general as stimulus duration was increased from 100 msec. to 400 msec., and as the duration became still longer (up to 2000 msec.) RT decreased again. This is in agreement with Wells's results with visual stimuli, but not with Wells's auditory data or with Froeberg's visual data.

In spite of the conflicting results obtained, it does seem as though there is a relationship between stimulus duration and RT. The most reasonable expectation is that the function is asymmetrical in nature, falling (RT becoming shorter) rapidly as the duration is lengthened from zero to some small time value, whereupon it rises more gradually to become asymptotic to some limit. The minimum of this curve is itself probably some function of stimulus intensity and length of foreperiod.

RT as a Function of the Onset and of the Cessation of the Stimulus

Another feature of the stimulus presentation which has occasioned some research, but about which little conclusive can be said, is the use of the onset and of the cessation of the stimulus as the signal for response. Both Holmes (74) and Jenkins (79) report shorter RT's to the cessation of light. On the other hand, Woodrow (160) found no differences in speed of simple reaction to stimulus onset or cessation for either auditory or visual stimuli. Wells (155) has shown that individual differences are very marked here, some *S*s reacting more quickly to one or the other. Woodworth (161) concludes that there is no difference between the two conditions, but the matter seems far from settled. Studies are required which control the subjective intensity of the stimulation at the time of onset and of cessation. The duration of the stimulus would appear to be an important variable differentially determining the effectiveness of each condition. The readiness of *S* might be expected to be different in accordance with the one to which he responds.

CENTRAL AND MOTOR FACTORS

Cerebral factors have not been studied in a way to make them applicable to this discussion, although a little work has been done (135). Studies of speed of nerve conduction have theoretical significance but do not allow for generalization since they deal with the application of the stimulus directly to the nerve fiber and by-pass both sensory and motor factors. Certain psychological condi-

tions which might be of importance,
e.g., personality factors (12, 151), the
influence of incentives and punish-
ments (80), etc., will not be dis-
cussed owing to the limited research
that has been done and the great
complexities of the results. A few
topics that might have been included
here have been singled out for the
next section, Special Factors.

Regarding the motor system, most
of the work that has been done on the
responding mechanism is irrelevant
in this context. Studies of the refrac-
tory phase (e.g., 142) are of interest
only in the case of successive quickly
elicited responses (39, 123). Most of
the relevant studies of this sort in-
volve more complex types of reaction
time measures.

RT as a Function of Age and Sex

The correlation of age and RT has
received some attention. Jones (84)
found an increase in speed of reaction
to sound in boys ranging from 11 to
14 years. According to this study no
further increases should be expected
beyond age 14. Atwell and Elbel (3),
however, report continued small in-
creases in speed up to 17, the oldest
age used in their study. Bellis (6),
in a study of ages ranging from 4 to
60 years, observed a general shorten-
ing of both visual and auditory RT's
until age 30, after which latencies be-
gan to grow longer. Even at 60,
however, he found that RT is still
faster than it is at 10. Miles (107),
using three kinds of response (finger
pressing, finger lifting, and right-foot
lifting), tested 100 adults between 25
and 87 years of age and found low
(0.25 to 0.55) but significant positive
correlations between age and speed of
reaction. Elliot and Louttit (38) also
report a low positive correlation. It
would appear, therefore, that RT
does covary with age. "Years old,"

however, may not be the best meas-
ure of the age factor.

Regarding sex differences, Elliot
and Louttit (38) in an investigation
of the braking reaction of men and
women in automobiles reported that
men react significantly more quickly
Bellis' (6) data, based on both audi-
tory and visual stimuli, also favor
males, especially in the age periods of
4–10 and 40–60. Seashore and Sea-
shore (132), in studying the speed of
various muscular responses (right
and left hands, right and left feet,
jaws), found men significantly faster,
especially after practice. The weight
of the evidence indicates that a sex
difference favoring men does exist,
although this conclusion is hardly
likely to end a perennial controversy.

RT as a Function of Preparatory Set

It is reasonable that RT will de-
pend on the degree to which S is
ready to respond. The use of a pre-
paratory signal has been shown to
yield faster RT's than the omission
of such a signal (163) or the unex-
pected presentation of a second sig-
nal requiring a second reaction (123).
Consequently, most experiments use
a ready signal.

This factor of readiness or *set* seems
to depend, among other things, on
the length of time between the warn-
ing or ready signal and the stimulus
to which the response is made, what
is known as the *foreperiod of reaction.*
Breitweiser (11) found definite indi-
vidual differences in the length of the
optimum foreperiod and reported a
range of optima between 1.0 and 4.0
sec. Telford (142), however, taking
repeated measurements of 29 Ss, ob-
tained conflicting results when he
found that the average RT increased
systematically from an optimum of
1.0 sec. to at least 4.0 sec. Telford's
data also indicated that as the fore-

period is reduced from the 1.0 sec. optimum to at least 0.5 sec., RT deteriorates markedly.

The study most frequently quoted with regard to the foreperiod is that of Woodrow (159). According to this study a 2.0-sec. foreperiod is optimum. In spite of the wide acceptance of this figure, there is a question as to the significance of Woodrow's results since the data were obtained from only three *S*s and, as they are reported, do not allow for an estimation of the standard errors of the means. Freeman and Kendall (54) estimated that if Woodrow's standard errors (which were not reported) were of the same order as those obtained in their study, which used four *S*s, the difference between the 2-sec. and 8-sec. intervals obtained by Woodrow would have fallen between the 5 and 10 per cent levels of confidence. Ordinarily this would be sufficient to accept the null hypothesis.

Early studies (85, 158) demonstrated that preparatory sets are primarily muscular. Livingston (95) has shown that the amount of muscular tension varies during the foreperiod. Freeman (52, 53) and Freeman and Kendall (54) have investigated the influence of the amount, the locus, and the time of induction of muscular tension during the foreperiod on the RT. Under conditions of heavy load (large induced muscular tension), a longer foreperiod was found to be optimum than under light-load conditions. The locus of the tension was also found to be a significant factor in determining the optimum preparatory interval. Perhaps their most interesting finding was that the optimum interval varied with the length of time prior to the response that the tension was induced. When the amount, locus, and time of induction of muscular tension

are all considered together, they (54) found that the optimum foreperiod ranged between 4 and 8 sec., depending on individual differences. As implied above, there is reason to believe that Woodrow's obtained optimum actually may be expressed best as a range between 2 and 8 sec.

Another factor thought to influence the readiness of the *S* to respond is whether he is set to respond to the stimulus ("sensory attitude") or whether he concentrates on the response ("muscular attitude"). Although the weight of the evidence appears to favor the latter, no controlled studies have been performed since Woodworth's (161) review. We are forced, therefore, to rest with the older, and for the most part less well-controlled, studies. These are thoroughly discussed by Woodworth.

The effect of instructions constitutes another great unknown which presumably determines the readiness of the *S* to respond. Perhaps the most serious obstacle to overcome in investigations of this factor is not the ambiguity of language but an inability to know and to control *S*'s self-instructions. Some promising work has been done, however, in which instructions were varied. Davis (33) compared the effect of instructions to respond with instructions to *not-respond*. Instructions to respond resulted in higher prestimulus tension levels in the responding arm, as measured by action potentials, and resulted in faster RT's. Davis also noted that instructions to respond were made even more effective by using more intense stimulation. This, however, would be predicted from the relationship between RT and stimulus intensity and may, therefore, be independent of the effects of instructions.

Moore (111) found that instruc-

tions have a greater influence on the variability of the RT than they do on the variability of the speed of movement of the responding part. When *S*s were instructed to respond as quickly as possible with the fastest possible movement, the speed of movement remained practically constant but the speed of reaction varied considerably.

It is clear that a great many factors influence the optimum foreperiod. Other variables which should be studied in this relation are the intensity and duration of the stimulus. Of these, duration would appear especially important since the durations of both ready signal and stimulus are confounded with the length of the foreperiod. No single value seems acceptable as *the optimum* since so many conditions are effective. Perhaps a more useful concept would be a range, the modal point of which would shift according to sensory modality, stimulus intensity and duration, nature of muscular tension, kind of instructions, etc. Presently available data indicate that such a range lies between approximately 1.5 and 8 sec.

RT as a Function of Body Position

One question of much interest is whether or not the RT varies according to the position of the body. This has especial relevance today in connection with many military and industrial problems. Reference to one such study has been found. Munnich (114) investigated the RT of several responses modeled after various aspects of an airplane pilot's task. These RT's were studied with *S* in six different bodily positions: (*a*) seated normally with the back of the chair making a 120° angle with the seat; (*b*) stomach downward; (*c*) head downward; (*d*) back downward; (*e*) on right side; (*f*) on left side. Un-

fortunately, the report was not available at the time of writing and the abstract does not describe the differences, if any, which were obtained between positions. One interesting result, however, is described. It was found that RT increased in general after any *change* in position, but as the new position was maintained the RT's tended to return to the values normal for it. The generality of this conclusion is offset by the finding that some *S*s improved their speed of reaction even after being changed to the most uncomfortable positions.

RT as a Function of the Responding Member

Of great theoretical and applied importance is the question of whether the RT varies with the body member which responds and with the type of motion involved. Regarding the latter, most of the responses which have been used have consisted of the simple release of a telegraph or similar key, or, less frequently, the depression of such a key. Other types of hand motion have also been used and, in addition, such other body parts as the eye, mouth (both movement of the jaws and the verbal response), arm, leg, foot, or the entire body. Although the types of response are usually as simple as possible and the methods of measurement usually some sort of chronoscopic device, both of these vary somewhat from one experiment to another. Discussions of the experimental techniques used and of the effects of these techniques may be found in several places (8, 11, 161). Whether the equivocal results obtained from studies of this type are really due to differences in technique is hard to say.

Woodworth (161), in reviewing this problem, presented a number of studies in which no differences were found between body members.

Baxter's (4) data support these results. On the other hand, Féré (44) found that the RT for the left hand was slower than that for the right, and that the sum of the RT's of each hand was larger when S tried to respond with both hands at once than when he moved them successively. Cattell (20) also observed that the RT for a particular task differs with the body member used, at least for the wrist, forearm, and shoulder. Hathaway (66) obtained results indicating that RT is longer for movement of the entire arm than it is for finger movement. Seashore and Seashore (132) reported that RT with the left hand was slower than with the right, and, similarly, the left foot was slower than the right foot. They also found that speed of reaction of the jaws was greater than either the hands or feet. Chernikoff and Taylor (23) presented data showing that when the stimulus is the sudden displacement of S's arm, his RT is shorter when he stops the movement than when he releases a telegraph key. Considering the contradictions in the various sets of data, it certainly seems too early to draw even a tentative conclusion about the role played, if any, by the kind of responding member. It is possible that the only real differences are due to differences in inertia of the responding musculatures.

Woodworth (161) suggested that complex movements produce longer RT's than do more simple ones. Not many studies were available to support this, but on the basis of those few that were, Woodworth made two specific hypotheses: (a) that guided (aimed) movements have longer RT's than those made freely; (b) that the RT is shorter for the initiation of a movement than it is for the stopping of it or for the changing of its direction.

No studies were found bearing directly on either of these hypotheses, but there are some results which are relevant to the general hypothesis that RT is somehow related to the response movement that follows it. Searle and Taylor (129) investigated the RT of corrective tracking movements made with a small knob to the displacement of a visual stimulus and found that this RT was not affected by the amount of knob movement required to make a correction. Brown and Slater-Hammel (13) reported that the RT involved in making discrete arm movements in the horizontal plane is independent of the length and direction of the movement. Henry (71) obtained data which indicate that RT's and movement times are not related. Except for the possible effect of the preparatory set, it is difficult to see why RT and consequent movement should be related. The results available indicate that they are not.

Other Central-Motor Factors

Studies of extraneous tension during the occurrence of the response appear to be of only incidental interest here. Action potentials taken from muscles not involved in the response are not related to the speed of reaction, according to Meyer (105, 106). Daniel (30), however, considered Meyer's conclusion not warranted by his data, and Henderson (69) reported a relationship between speed of reaction of one arm and action potentials taken from the nonparticipating arm.

Of greater interest in this context is the finding that lowered muscular tension produces longer RT's (86, 87, 147). This finding appears to be related to the problem of the preparatory set and to the effects of practice. A few writers (e.g., 69) have reported increased speeds with practice. It is

possible that these increases may be due, not to the effect of learning on the RT itself, but to the effect of learning on the preparatory interval. This latter effect may consist of the learning of an optimal anticipatory muscular tension.

The RT exhibits considerable variability among individuals (41, 60, 92, 161) but is, nevertheless, reasonably consistent. Reported reliability coefficients range between 0.83 and 0.92 (41, 42, 59, 115, 130, 131). RT measures also show considerable variability among experimental studies. Both types of variability may be due, in part, to the points on the tremor cycle at which the stimulus is presented (145). In spite of variability, relatively high intercorrelations among different kinds of RT tests have been obtained (42), which suggests a common factor present in the various tests, and, according to Seashore, Buxton, and McCollom (130), such a factor can be analyzed. Further evidence supporting the notion of an RT factor comes from studies showing no relationship between RT and intelligence (42, 133), substitution (133), card sorting (133), discrimination reaction time (42, 130), pursuit rotor performance (42, 130), and little or no relationship to tapping tests (130). Slocombe and Brakeman (134) have also suggested that RT measures yield a group factor and, in addition, that this factor may be used to discriminate, at least among motormen, between good and poor accident risks.

SPECIAL FACTORS

Certain conditions, such as exposure to extreme climatic conditions, the effects of drugs, prolonged effort, starvation, deprivation of sleep for long periods, etc., are singled out here because of the unique interest which they hold for many investigators. Some, if not all, of these could have been included in the previous section.

RT as a Function of Prolonged Readiness

Mackworth (102) in a series of studies of military personnel under conditions of prolonged vigilance found that the RT to the double-jumps of a clock hand increased sharply after one-half hour of watching. This increase could be prevented by the use of Benzedrine or by informing S of the results of the test.

In a related study Kennedy and Travis (87) measured the action potentials and the RT to combined light and sound stimuli under long monotonous conditions. Their results indicate that the frequency of action potentials decreases and the RT to irregularly presented stimulation becomes longer as the testing period is lengthened. After falling asleep, one stimulus presentation was sufficient to awaken S, the RT on this occasion being considerably longer, as might be expected. However, the RT approached its normal level rapidly with subsequent presentations.

RT as a Function of Certain Common Drugs

The RT has been used as the "I told you so" of all those who would restrict the use of the common drugs. In spite of this, the effects of these drugs on RT are not always clear when the conditions of the experiment are carefully considered. Some of the experimental problems are discussed by Hull (77) and more recently by Gray and Trowbridge (59) and by Miller (109). No real attempt will be made here to evaluate this type of effect but merely to point out a few representative papers.

With respect to coffee drinking, Hawk (68) and Gilliland and Nelson 58 claimed lengthened RT's. On the other hand, there are studies such as that of Thornton *et al.* (143) in which no effect was claimed. Cigarette smoking appears to decrease the variability of visual RT's according to both Hull (77) and Fay (43), but has no other reliable effect. Alcohol, on the other hand, both increases variability and lengthens visual (108, 150) and auditory (150) RT's. Benzedrine appears to have little or no effect on the auditory RT (143). Mackworth (102), however, did find that Benzedrine tended to offset decrements produced by prolonged vigilance. Aspirin has no effect on either visual or auditory RT's (31). A 30 per cent saturation of carbon monoxide is required to produce an increase in visual RT (49). Finally, morphine has the effect of first shortening and then lengthening RT except when taken in large doses, in which case only the latter effect occurs (101).

RT as a Function of Temperature

A number of studies (e.g., 75, 110, 157) have been done to determine the effects of climatic stress, temperature in particular, on the RT. The general result of these studies is that ambient temperatures between a range of −50°F. and 117°F. have little or no effect on either RT or more complex reaction times. This conclusion was reached by Forlano, Barmack, and Coakley (50) after a careful review of the effects of ambient and body temperatures on RT. Most of the studies available for evaluation, however, are distinguished by the degree with which several main variables are confounded in one experiment, and consequently are given to difficulty of interpretation. Such a conclusion,

therefore, should not be accepted as firmly established.

RT's have also had a little attention with regard to skin temperatures. Craik and Macpherson (29) report that cooling of the hand with which the response is made may increase RT by 10–15 per cent. This conclusion is not reasonable on the basis of their study since an increase of this much turns out to be an increase of 0.02 to 0.06 sec., a change which has little significance in terms of the likely error of measurement and the size of the sample (two Ss).

A few RT studies have been done with body temperature and/or time of day as the independent variable. In general, these studies (89, 90, 104) suggest that RT exhibits a slight diurnal variation, but with large individual differences. The data may also be interpreted to indicate that RT is a function of body temperature and is only spuriously correlated with time of day (50).

Effects of Sleep Conditions on RT

Most of the studies in which the stimulus for the RT has been presented during sleep have been related to the problem of determining the onset or the depth of sleep. Since the detection of both of these conditions still lacks other independent criteria, studies of this sort have little validity with respect to generalizations about the effect of sleeping on the RT. A discussion of this problem may be found in Kleitman (88).

Mullin and Kleitman (113) used a criterion of sleep onset which was independent of the RT and then investigated the change in a verbal, auditory RT for periods of time following onset. In this experiment, onset of sleep was considered to be established when *S* released a piece of paper which he held in his hand.

Using both normal and feebleminded adults and normal children as Ss, Mullin and Kleitman found a slow increase in the verbal RT for the first 25 min. after onset, followed by a stable period of long RT's, and then by a third period of rapid shortening of the RT. This cycle, which was sigmoid in nature, was completed during the first hour after onset and did not reappear during the rest of the night.

Considerably more work has been done concerning the effect of lack of sleep on RT. Although most of the studies were concerned primarily with more complex reaction times, enough have been done with the RT to allow some evaluation.

As long ago as 1896 Patrick and Gilbert (116) tested the auditory and visual RT's of three Ss deprived of sleep for 90 hours and reported a marked slowing of speed. Later, Robinson and Herrman (126), in a more rigorous experiment, found that experimental insomnia produced no consistent effect on RT. With the exception of one or two studies reviewed by Kleitman (88), this latter result has marked the literature ever since, not only with respect to RT measures but practically all measures taken. Part of the difficulty might be blamed on the small samples usually used. On the other hand, large samples are administratively extremely difficult to achieve in this type of investigation. Lee and Kleitman (93), for example, tested the RT of only one S following 114 hours of sleep deprivation. This one S showed no effect. Cooperman, Mullin, and Kleitman (28) report no change in auditory RT for six Ss after 60 hours of privation, as did Tyler (149) after 24–114 hours of sleep loss.

Edwards (36) reported a study involving relatively large samples. He compared the auditory RT of 17 Ss deprived of sleep for 100 hours with that of ten control Ss. His data support the studies already cited since he found no significant differences between the groups. It seems, therefore, that although extremely small samples may contribute to the variability of results, they cannot be considered the major reason for the failure of more recent studies to confirm the results of Patrick and Gilbert.

Most writers, when discussing the general failure to show that sleep loss influences RT, invoke "compensation" on the part of S as an explanatory concept. What is usually meant is that S achieves the same performance under stress as he does when not under stress by expending greater effort. Granting the looseness of this argument, the possibility of testing it appears available. If one is willing to accept greater muscular tension during the foreperiod and during the RT interval as a measure of greater effort, it should be relatively easy to determine by measuring tension during these intervals if compensation is really a useful concept.

Other Special Effects

One experiment was found in which the effect on the RT of radial acceleration was studied. Canfield, Comrey, and Wilson (16) investigated the effects of positive acceleration forces of 1, 3, and 5 g on the RT's to light and also to sound. Both kinds of RT were found to lengthen significantly with increased acceleration. Moreover, the slowing of reaction observed was not due to decrements in sensory functioning as the Ss were neither "blacked out" nor "greyed out" and, in addition, the Ss did not notice any change in sensory functioning.

In a different type of study, Tuttle et al. (148) reported that omission

of breakfast increased the visual RT of their five Ss. In a more careful nutrition study involving restriction of vitamin B, Brožek *et al.* (12) found that partial restriction of vitamin B intake for 23 days had no effect on the auditory RT of their eight Ss. Only prolonged and severe deprivation produced significant decrements.

McFarland (97, 98, 99, 100) has shown that RT is affected by altitude (anoxia), but not significantly before at least 20,000 ft. Variability of RT, however, begins to increase at 2000–3000 ft. before the average RT shows a decrement. McFarland (97) has also reported that people who live on mountains at extreme altitudes have longer and more variable RT's than people living at sea level in the same region.

Summary and Conclusions

Factors other than those discussed would appear to be important, but have not been studied in relation to the RT. For example, it might be expected that RT would decrease as a function of speed of travel or stimulation across the retina or the skin. The condition of the receptor and of the responding member are still among the conditions which have received insufficient attention. In spite of a long history of research, the effect of the sensory system is still an open question.

Many of our notions about the RT depend upon what are now classical studies, studies which did not have the advantage of modern statistical procedures. The question of mono- versus bisensory stimulation falls in this class. Research is also required to ascertain the effect of stimulus duration and the interaction between duration and intensity. Related to this is the problem of area or number of receptors stimulated, and

this too is in need of further experimental investigation.

On the positive side, intensity functions are relatively well established. The last 20 years have yielded a considerable amount of information about the effects of age and sex. They have also produced more definitive information about the role of the responding member, and about the factors involved in the foreperiod, although a great deal more research must yet be performed before the latter is really well understood.

The incentive to study the effects of sleep loss, drugs, temperature, altitude, acceleration, vigilance, etc. to a large extent has come from applied areas. Few, if any, safe generalizations are yet available. However, these topics will undoubtedly continue to receive a great deal of attention since many of them constitute important military problems.

All things considered, the last two decades have been quite productive, not of mathematically expressed empirical laws, but of useful nonmathematical generalizations and of clearer definition of the problems than was available before. The present status of the RT study may best be evaluated in terms of the generalizations which it has yielded. In the opinion of the writer the following generalizations appear to have been reasonably well established.

1. There is a positive correlation between the visual and the auditory RT.

2. Simultaneous stimulation of more than one sense modality produces faster RT's than stimulation of just one. On the other hand, successive stimulation of different senses produces slower RT's than stimulation of a single sensory channel.

3. For visual and thermal RT's the greater the extent of the stimulus in

space, i.e. the greater the number of receptors stimulated, the faster the speed of reaction up to some limit.

4. Under daylight or illuminated conditions the visual RT becomes longer the greater the distance of stimulation from the fovea.

5. In the case of each receptor system, RT is a negatively accelerated decreasing function of intensity up to some maximum intensity value after which RT either becomes suddenly lengthened, the function at this point being discontinuous, or asymptotic to a physiological limit.

6. RT is a slowly falling growth function of chronological age until about 30 years after which it is a slowly rising function.

7. In general the RT of the human male is faster than that of the female.

8. The optimum foreperiod of RT may be thought of as lying in a range between approximately 1.5 and 8.0 sec. Its position in this range is determined by a large number of factors including the duration and intensity of the warning signal and of the stimulus, and the amount, locus, and time of production of muscular tension.

9. RT is not related to the length, direction, or speed of movement of the responding member.

10. Under vigilance conditions, the longer the period during which S must respond, the longer the RT.

REFERENCES

1. ABEL, THEODORA M. The influence of visual and auditory patterns on tactual recognition. *Amer. J. Psychol.*, 1934, 46, 443–447.
2. ARNOLD, D. C., & TINKER, M. A. The fixation pause of the eyes. *J. exp. Psychol.*, 1939, 25, 271–280.
3. ATWELL, W. O., & ELBEL, E. R. Reaction time of male high school students in 14–17 year age groups. *Res. Quart. Amer. Ass. Hlth.*, 1948, 19, 22–29.
4. BAXTER, B. A. A study of reaction time using factorial design. *J. exp. Psychol.*, 1942, 31, 430–437.
5. BAXTER, B., & TRAVIS, R. C. The reaction time to vestibular stimuli. *J. exp. Psychol.*, 1938, 22, 277–282.
6. BELLIS, C. J. Reaction time and chronological age. *Proc. Soc. exp. Biol. Med.*, 1933, 30, 801–803.
7. BERGER, G. O. Ueber den Einfluss der Reizstärke auf die Dauer ein facher psychiser Vorgänge mit besonderer Rücksicht auf Lichtreize. *Phil. Stud.*, 1886, 3, 38–93.
8. BILLS, A. G. Studying motor functions and efficiency. In T. G. Andrews, (Ed.), *Methods of psychology.* New York: Wiley, 1947.
9. BLISS, C. B. Investigations in reaction time and attention. *Stud. Yale psychol. Lab.*, 1893, 1, 1–55.
10. BORING, E. G. *Physical dimensions of consciousness.* New York: Century, 1933.
11. BREITWIESER, J.V. Attention and movement in reaction time. *Arch. Psychol.*, 1911, No. 4.
12. BROWER, D., & SANDS, H. Relations between reaction time and personal adjustment as measured by the Bell Adjustment Inventory. *J. gen. Psychol.*, 1948, 38, 229–233.
13. BROWN, J. S., & SLATER-HAMMEL, H. T. Discrete movements in the horizontal plane as a function of their length and direction. *J. exp. Psychol.*, 1949, 39, 84–95.
14. BROŽEK, J., GUETZKOW, H., & KEYS, A. A. A study of the personality of normal young men maintained on restricted intakes of vitamin B complex. *Psychosom. Med.*, 1946, 8, 98–109.
15. CANFIELD, A. A., JR. The influence of positive g on reaction time. *Amer. Psychologist*, 1950, 5, 362. (Abstract)
16. CANFIELD, A. A., COMREY, A. L., & WILSON, R. C. A study of reaction time to light and sound as related to increased positive radial acceleration. *J. Aviat. Med.*, 1949, 20, 350–355.
17. CARLSON, W. R., DRIVER, R. C., & PRESTON, M. G. Judgment times for the method of constant stimuli. *J. exp. Psychol.*, 1934, 17, 113–118.
18. CASSEL, E. E., & DALLENBACH, K. M. The effect of auditory distraction upon the sensory reaction. *Amer. J. Psychol.*, 1918, 29, 129–143.
19. CATTELL, J. McK. The influence of the

intensity of the stimulus on the length of the reaction time. *Brain*, 1886, 8, 512–515.

20. CATTELL, J. McK. On reaction time and velocity of the nerve impulse. *Nat. Acad. sci. Mem.*, 1893, 7, 410.

21. CHERNIKOFF, R., & BROGDEN, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, 42, 357–364.

22. CHERNIKOFF, R., GREGG, L. W., & BROGDEN, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, 43, 123–128.

23. CHERNIKOFF, R., & TAYLOR, F. V. Reaction time to kinesthetic stimulation resulting from sudden arm displacement. *J. exp. Psychol.*, 1952, 43, 1–8.

24. CHOCHOLLE, R. Etude de la psychophysiologie de l'audition par la méthode des temps de réaction. *Année Psychol.*, 1948, 45–56, 90–131.

25. CHOCHOLLE, R. Quelques remarques sur les variations et la variabilité des temps de réaction auditifs. *J. Psychol. norm. path.*, 1948, 41, 345–358.

26. COBB, P. W. Some experiments on speed of vision. *Trans. illum. engng Soc.*, 1924, 19, 150–175.

27. COERMANN, R. Untersuchungen über die Einwirkung von Schwingungen auf den menschlichen Organismus. *Industr. Psychotech.*, 1939, 16, 169–206.

28. COOPERMAN, N. R., MULLIN, F. J., & KLEITMAN, N. Studies on the physiology of sleep. XI. Further observations on the effects of prolonged sleeplessness. *Amer. J. Physiol.*, 1934, 107, 589–593.

29. CRAIK, K. J. W., & MACPHERSON, S. J. Effects of cold upon hand movement and reaction time. Report Comm. Armoured Vehicles, MPRCBPC 43/196, March 13, 1943.

30. DANIEL, R. S. Some observations on Meyer's study of reaction time and muscle tension. *J. exp. Psychol.*, 1949, 39, 896–898.

31. DAVIS, R. C. The effects of analgesic dosage of aspirin (acetyl salicylic acid) on some mental and motor performances. *J. appl. Psychol.*, 1936, 20, 481–487.

32. DAVIS, R. C. Motor components of responses to auditory stimuli: the relation of stimulus intensity and instruc-

tions to respond. *Amer. Psychologist*, 1947, 2, 308. (Abstract)

33. DAVIS, R. C. Motor effects of strong auditory stimuli. *J. exp. Psychol.*, 1948, 38, 257–275.

34. DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, 40, 107–120.

35. EDES, B., & DALLENBACH, K. M. The adaptation of pain aroused by cold. *Amer. J. Psychol.*, 1936, 48, 307–315.

36. EDWARDS, A. S. Effects of the loss of one hundred hours of sleep. *Amer. J. Psychol.*, 1941, 54, 80–91.

37. EDWARDS, W. Recent research on pain perception. *Psychol. Bull.*, 1950, 47, 449–474.

38. ELLIOT, F. R., & LOUTTIT, C. M. Auto braking reaction times to visual vs. auditory warning signals. *Proc. Ind. Acad. Sci.*, 1948, 47, 220–225.

39. ELLSON, D. G., & HILL, H. The interaction of responses to step function stimuli. I. Opposed steps of constant amplitude. *USAF, Aero Med. Lab.*, *Wright-Patterson AFB*, MCREXD 694-28, 1948.

40. ESCHER-DESRIVIÈRES, J. Variations des temps de réactions psychomotrices visuelles en fonction de l'éclairement en lumière blanche et colorée. *C. R. Acad. Sci., Paris*, 1939, 208, 1751–1753.

41. FARMER, E., & CHAMBERS, E. G. A psychological study of individual differences in accident rates. *Med. Res. Coun., Industr. Hlth Res. Bd.*, Report No. 38, 1926, London, Eng.

42. FARNSWORTH, P. R., SEASHORE, R. H., & TINKER, M. A. Speed in simple and serial action as related to performance in certain "intelligence" tests. *J. genet. Psychol.*, 1927, 34, 537–551.

43. FAY, P. J. The effect of cigarette smoking on simple and choice reaction time to colored lights. *J. exp. Psychol.*, 1936, 19, 592–603.

44. FÉRÉ, F. L. L'energie et la vitesse des mouvements volontaires. *Rev. Phil.*, 1889, 28, 36–69.

45. FERREE, C. E., & RAND, G. Intensity of light and speed of vision studied with special reference to industrial situations. Part I. *Trans. illum. engng Soc.*, 1927, 22, 79–110.

46. FINAN, J. L., FINAN, S. C., & HARTSON, L. D. A review of representative tests used for the quantitative measure-

ments of behavior-decrement under conditions related to aircraft flight. Dayton, O.: U. S. Air Material Command, Wright-Patterson Air Force Base, 1949. iv, 230 p. (*USAF Tech. Rep.* No. 5830.)

47. FINCH, G. Review of muscle activity and action potentials as they are related to movement. (AAF AMC Aero Med. Lab. Memo Rep. TSEAA 694-2E, 1947; Publ. Bd. No. M 81423.) Washington, D. C.: U. S. Dep. Commerce, 1947.

48. FORBES, G. The effect of certain variables on visual and auditory reaction times. *J. exp. Psychol.*, 1945, **35**, 153–162.

49. FORBES, W. H., DILL, D. B., DE SILVA, H., & VANDEVENTER, F. M. The influence of moderate carbon monoxide poisoning upon the ability to drive automobiles. *J. indust. Hyg. Toxicol.*, 1937, **19**, 598–608.

50. FORLANO, G., BARMACK, J. E., & COAKLEY, J. D. The effect of ambient and body temperatures upon reaction time. *Special Devices Center*, Report No. 151-1-13, 1948.

51. FRANKLIN, J. C., & BROŽEK, J. The relation between distribution of practice and learning efficiency in psychomotor performance. *J. exp. Psychol.*, 1947, **37**, 16–24.

52. FREEMAN, G. L. The optimal locus of anticipatory tensions in muscular work. *J. exp. Psychol.*, 1937, **21**, 554–564.

53. FREEMAN, G. L. The optimal muscular tensions for various performances. *Amer. J. Psychol.*, 1938, **51**, 146–150.

54. FREEMAN, G. L., & KENDALL, W. E. The effect upon reaction time of muscular tension induced at various preparatory conditions. *J. exp. Psychol.*, 1940, **27**, 136–148.

55. FROEBERG, S. The relation between the magnitude of stimulus and the time of reaction. *Arch. Psychol.*, 1907, **16**, No. 8, 1–38.

56. FROELICH, F. W. *Die Empfindungszeit.* (Ed. 1) Jena: Fischer, 1929.

57. GALIFRET, Y., & PIÉRON, H. Vitesse de réaction et intensité de sensation. Donnees experimentales sur le problème d'une courbe sigmoid des vitesses. *Année Psychol.*, 1951, **51**, 1–16.

58. GILLILAND, A. R., & NELSON, D. The effects of coffee on certain mental and physiological functions. *J. gen. Psychol.*, 1939, **21**, 339–348.

59. GRAY, M. G., & TROWBRIDGE, E E. Methods for investigating the effects of drugs on psychological function. *Psychol. Rec.*, 1942, **5**, 127–148.

60. GREENSHIELDS, B. D. Reaction time in automobile driving. *J. appl. Psychol.*, 1936, **20**, 353–358.

61. GREGG, E. C., JR. Physical basis of pain threshold measurements in man. *J. appl. Physiol.*, 1951, **4**, 351–363.

62. GREGG, L. W., & BROGDEN, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, **43**, 329–337.

63. GREGG, L. W., & BROGDEN, W. J. The relation between reaction time and the duration of the auditory stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 389–395.

64. GUILFORD, J. P., & EWART, E. Reaction time during distraction as an indication of attention-value. *Amer. J. Psychol.*, 1940, **53**, 554–563.

65. HAMEL, I. A. A study and analysis of the conditioned reflex. *Psychol. Monogr.*, 1919, **27**, No. 1 (Whole No. 118).

66. HATHAWAY, S. R. An action potential study of neuromuscular relations. *J. exp. Psychol.*, 1935, **11**, 285–298.

67. HATHAWAY, S. R., & SISSON, E. D. The time relations of the events in quick voluntary movements. *Psychol. Bull.*, 1935, **32**, 721–722.

68. HAWK, P. B. A study of the physiological and psychological reactions of the human organism to coffee drinking. *Amer. J. Physiol.*, 1929, **90**, 380–381.

69. HENDERSON, R. L. Remote action potentials at the moment of response in a simple reaction-time situation. *J. exp. Psychol.*, 1952, **44**, 238–241.

70. HENMON, V. A. C., & WELLS, F. L. Concerning individual differences in reaction times. *Psychol. Rev.*, 1914, **21**, 153–156.

71. HENRY, F. M. Independence of reaction and movement times and equivalence of sensory motivators of faster response. *Res. Quart. Amer. Ass. Hlth*, 1952, **23**, 43–53.

72. HICK, W. E. Reaction time for the amendment of a response. *Quart. J. exp. Psychol.*, 1949, **1**, 175–179.

73. HILDEN, A. H. An action current study of the conditioned hand withdrawal. *Psychol. Monogr.*, 1937, **49**, No. 1 (Whole No. 217), 173–204.

74. HOLMES, J. L. Reaction time to light as conditioned by wave-length and in-

tensity. Unpublished doctor's dissertation, Columbia Univer., 1923.

75. HORVATH, S. M., & FREEDMAN, A. The influence of cold upon the efficiency of man. *J. Aviat. Med.*, 1947, **18**, 158–164.

76. HOVLAND, C. I. The influence of adaptation illumination upon visual reaction time. *J. gen. Psychol.*, 1936, **14**, 346–359.

77. HULL, C. L. The influence of tobacco smoking on mental and motor efficiency. *Psychol. Monogr.*, 1924, **33**, No. 3, (Whole No. 150), 1–160.

78. HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

79. JENKINS, T. N. Facilitation and inhibition. *Arch. Psychol.*, 1926, No. 86, 1–56.

80. JOHANSON, A. M. The influence of incentive and punishment upon reaction-time. *Arch. Psychol.*, 1922, No. 54, 1–52.

81. JOHNSON, H. M. The influence of the distribution of brightnesses over the visual field on the time required for discriminative responses to visual stimuli. *Psychobiol.*, 1918, **1**, 459–494.

82. JOHNSON, H. M. Reaction time measurements. *Psychol. Bull.*, 1923, **20**, 562–589.

83. JONES, B. F., FLINN, R. H., HAMMOND, E. C., *et al.* Fatigue and hours of service of Interstate Truck Drivers. *Pub. Hlth Bull.*, 1941, No. 265, Fed. Sec. Agency, U. S. Pub. Hlth. Ser., Wash., D. C.

84. JONES, H. E. *Motor performance and growth.* Berkeley: Univer. of California Press, 1949.

85. JUDD, C. H., MCALLESTER, C. H., & STEELE, W. M. Analysis of reaction movements. *Psychol. Monogr.*, 1904, **7**, No. 1, (Whole No. 29), 141–184.

86. KENNEDY, J. L., & TRAVIS, R. C. Prediction and control of alertness. II. Continuous tracking. *J. comp. physiol. Psychol.*, 1947, **41**, 203–210.

87. KENNEDY, J. L., & TRAVIS, R. C. Prediction of speed of performance by muscle action potentials. *Science*, 1947, **105**, 410–411.

88. KLEITMAN, N. *Sleep and wakefulness.* Chicago: Univer. of Chicago Press, 1939.

89. KLEITMAN, N., & JACKSON, O. P. Body temperature and performance under different routines. *J. appl. Physiol.*, 1950, **3**, 304–328.

90. KLEITMAN, N., TITELBAUM, S., & FEIVESON, P. The effect of body temperature on reaction time. *Amer. J. Physiol.*, 1938, **121**, 495–501.

91. LANDAHL, H. D. Contributions to the mathematical biophysics of the central nervous system. *Bull. math. Biophysics*, 1939, **1**, 95–118.

92. LANIER, L. H. The interrelations of speed and reaction measurements. *J. exp. Psychol.*, 1934, **17**, 371–399.

93. LEE, M. A. M., & KLEITMAN, N. Studies on the physiology of sleep. II. Attempts to demonstrate functional changes in the nervous system during experimental insomnia. *Amer. J. Physiol.*, 1923, **67**, 141–151.

94. LEMMON, V. W., & GEISINGER, S. M. Reaction time to retinal stimulation under light and dark adaptation. *Amer. J. Psychol.*, 1936, **48**, 140–142.

95. LIVINGSTON, W. A. Action potential measurements from the arm in the foreperiod of reaction time to visual stimuli. *Proc. Ind. Acad. Sci.*, 1946, **55**, 170. (Abstract)

96. LUCKIESH, M. *Light, vision, and seeing.* New York: Van Nostrand, 1944.

97. MCFARLAND, R. A. The psychological effects of oxygen deprivation (anoxemia) on human behavior. *Arch. Psychol.*, 1932, No. 145, 1–135.

98. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. I. The effect of rapid ascents by aeroplane and train. *J. comp. Psychol.*, 1937, **23**, 191–225.

99. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. II. Sensory and motor responses during acclimatization. *J. comp. Psychol.*, 1937, **23**, 227–258.

100. MCFARLAND, R. A. Psycho-physiological studies at high altitude in the Andes. IV. Sensory and circulatory responses of the Andean residents at 17,500 ft. *J. comp. Psychol.*, 1937, **24**, 189–220.

101. MACHT, D. I., & ISAACS, S. Action of some opium alkaloids on the psychological reaction time. *Psychobiol.*, 1917, **1**, 19–32.

102. MACKWORTH, N. H. *Researches on the measurement of human performance.* London: His Majesty's Stationery Office, 1950. (Med. Res. Coun., Special Rep. Ser., No. 268.)

103. MALLORY, E. B. The recognition of relatively simple sensory experiences. *Amer. J. Psychol.*, 1943, **46**, 120–131.

104. MARSH, H. D. The diurnal course of efficiency. *Arch. Phil. Psychol. Sci. Methods*, 1906, No. 7.

105. MEYER, H. D. Reaction time as related to tensions in muscles not essential in the reaction. *J. exp. Psychol.*, 1949, **39**, 96–113.

106. MEYER, H. D. Some remarks concerning Daniel's observations. *J. exp. Psychol.*, 1949, **39**, 898–900.

107. MILES, W. R. Correlation of reaction and coordination speed with age in adults. *Amer. J. Psychol.*, 1931, **43**, 377–391.

108. MILES, W. R. *Alcohol and human efficiency.* Washington, D. C.: Carnegie Inst., 1936.

109. MILLER, L. C. A critique of analgesic testing methods. *Ann. N. Y. Acad. Sci.*, 1948, **51**, 34–50.

110. MITCHELL, H. H., GLICKMAN, U., LAMBERT, E. H., KEETON, R. W., & FAHNESTOCK, M. K. The tolerance of man to the cold as affected by dietary modification: carbohydrate versus fat and the effect on the frequency of meals. *Amer. J. Physiol.*, 1946, **146**, 84–96.

111. MOORE, T. V. A study of reaction time and movement. *Psychol. Rev. Monogr. Suppl.*, 1904, **6**, No. 1 (Whole No. 24).

112. MOWRER, O. H. Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.*, 1940, **52**, No. 2 (Whole No. 233).

113. MULLIN, F. J., & KLEITMAN, N. Variations in threshold of auditory stimuli necessary to awaken the sleeper. *Amer. J. Physiol.*, 1938, **123**, 477–481.

114. MÜNNICH, K. Die Reaktionsleistung in Abhängigkeit von der Körperlage. *Industr. Psychotech.*, 1940, **17**, 49–83. (See *Psychol. Abstr.*, 1947, No. 1649.)

115. MUSCIO, B. On the relation of fatigue and accuracy to speed and duration of work. *Med. Res. Coun., Industr. Hlth Res. Bd.*, 1922, Report No. 19-B, London, Eng.

116. PATRICK, G. T. W., & GILBERT, J. A. On the effects of loss of sleep. *Psychol. Rev.*, 1896, **3**, 469–483.

117. PATTLE, R. E., & WEDDELL, G. Observations on electrical stimulation of pain fibres in an exposed human sensory nerve. *J. Neurophysiol.*, 1948, **11**, 93–98.

118. PIÉRON, H. Nouvelles recherches sus l'analyse du temps de latence senr sorielle et sur la loi qui relie le temp à l'intensitié d'excitation. *Année Psychol.*, 1920, **22**, 58–142.

119. PIÉRON, H. Recherches expérimentales sur la marge de variation du temps de latence de la sensation lumineuse par une méthode de masquage. *Année Psychol.*, 1926, **26**, 1–30.

120. PIÉRON, H. *The sensations: their functions, processes and mechanism.* (Trans. by M. H. Pierenne & B. C. Abbott.) New Haven: Yale Univer. Press, 1952.

121. POFFENBERGER, A. T. Reaction time to retinal stimulation, with special reference to the time lost in conduction through nerve centers. *Arch. Psychol.*, 1912, No. 23, 1–73.

122. POSTMAN, L., & KAPLAN, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.*, 1947, **37**, 136–145.

123. POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99–112.

124. RASHEVSKY, N. *Advances and applications of mathematical biology.* Chicago: Univer. of Chicago Press, 1940.

125. ROBINSON, E. S. Work of the integrated organism. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934.

126. ROBINSON, E. S., & HERMANN, S. O. Effects of loss of sleep. *J. exp. Psychol.*, 1932, **15**, 19–32.

127. ROOS, J. The latent period of skeletal muscle. *J. Physiol.*, 1932, **74**, 17–33.

128. SALTZMAN, I. J., & GARNER, W. R. Reaction time as a measure of span of attention. *J. Psychol.*, 1948, **25**, 227–241.

129. SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.

130. SEASHORE, R. H., BUXTON, C. E., & McCOLLOM, I. N. Multiple-factor analysis of fine motor skills. *Amer. J. Psychol.*, 1940, **53**, 251–259.

131. SEASHORE, R. H., STARMAN, R., KENDALL, W. E., & HELMICK, J. S. Group factors in simple and discriminative reaction times. *J. exp. Psychol.*, 1941, **29**, 346–349.

132. SEASHORE, S. H., & SEASHORE, R. H. Individual differences in simple auditory reaction times of hands, feet, and jaws. *J. exp. Psychol.*, 1941, **29**, 342–345.

133. SISK, T. K. The interrelations of speed in simple and complex responses.

Peabody Coll. Contr. Educ., 1926, No. 23.

144 SLOCOMBE, C. S., & BRAKEMAN, E. E. Psychological tests and accident proneness. *Brit. J. Psychol.*, 1930, 21, 29–38.

135. SMITH, K. U. The functions of the intercortical neurones in sensorimotor coordination and thinking in man. *Science*, 1947, 105, 234–235.

136. SMITH, W. M. Sensitivity to apparent movement in depth as a function of stimulus dimensionality. *J. exp. Psychol.*, 1952, 43, 149–155.

137. STEINMAN, A. R. Reaction time to change compared with other psychophysical methods. *Arch. Psychol.*, New York, 1944, No. 292, 34–60.

138. STEINMAN, A., & VENIAR, S. Simple reaction time to change as a substitute for the disjunctive reaction. *J. exp. Psychol.*, 1944, 34, 152–158.

139. STONE, L. J., & DALLENBACH, K. M. Adaptation to the pain of radiant heat. *Amer. J. Psychol.*, 1934, 46, 229–242.

140. STRUGHOLD, H. The human time factor in flight. The latent period of optical perception and its significance in high speed flying. *J. Aviat. Med.*, 1949, 20, 300–307.

141. STRUGHOLD, H. The human time factor in flight: II. Chains of latencies in vision. *J. Aviat. Med.*, 1951, 22, 100–108.

142. TELFORD, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, 14, 1–36.

143. THORNTON, G. R., HOLCK, H. G. O., & SMITH, E. L. The effect of benzedrine and caffeine upon performance in certain psychomotor tasks. *J. abnorm. soc. Psychol.*, 1939, 34, 96–113.

144. THURSTONE, L. L. Psychophysical methods. In T. G. Andrews (Ed.), *Methods of psychology*, New York: Wiley, 1947.

145. TIFFIN, J., & WESTHAFER, F. L. The relation between reaction time and temporal location of the stimulus on tremor cycle. *J. exp. Psychol.*, 1940, 27, 318–324.

146. TODD, J. W. Reaction to multiple stimuli. *Arch. Psychol.*, 1912, No. 25, 1–65.

147. TRAVIS, R. C., & KENNEDY, J. L. Prediction and automatic control of alertness. I. Control of lookout alertness. *J. comp. physiol. Psychol.*, 1947, 40, 457–461.

148. TUTTLE, W. W., WILSON, M., & DAUM, K. Effect of altered breakfast habits on physiologic response. *J. appl. Physiol.*, 1949, 1, 545–559.

149. TYLER, D. B. The effect of amphetamine sulfate and some barbiturates on the fatigue produced by prolonged wakefulness. *Amer. J. Physiol.*, 1947, 150, 253–262.

150. VARÉ, P. Influence de l'alcool sur les réactions psychometrices. *C. R. Soc. Biol.*, 1932, 11, 70–72.

151. VERBILLE, E. The effect of emotional and motivational sets on the perception of incomplete pictures. *J. genet. Psychol.*, 1946, 69, 133–145.

152. VINCE, M. A. Corrective movements in a pursuit task. *Quart. J. exp. Psychol.*, 1948, 1, 85–103.

153. WELLS, F. L., KELLEY, C. M., & MURPHY, G. Comparative simple reactions to light and sound. *J. exp. Psychol.*, 1921, 4, 57–62.

154. WELLS, F. L., KELLEY, C. M., & MURPHY, G. On attention and simple reaction. *J. exp. Psychol.*, 1921, 4, 391–398.

155. WELLS, G. R. The influence of stimulus duration on reaction time. *Psychol. Monogr.*, 1913, 15, No. 5 (Whole No. 66).

156. WERTHEIMER, M. A single-trial technique for measuring the threshold of pain by thermal radiation. *Amer. J. Psychol.*, 1952, 65, 297–298.

157. WILLIAMS, C. C., & KITCHING, J. A. The effects of cold on human performance. I. Reaction time. 1942, *Misc. Canad. Aviat. Rep.*, No. 81-A.

158. WILLIAMS, R. D. Experimental analysis of forms of reaction movement. *Psychol. Monogr.*, 1914, 17, No. 4 (Whole No. 75), 55–155.

159. WOODROW, H. The measurement of attention. *Psychol. Monogr.*, 1914, 17, No. 5 (Whole No. 76).

160. WOODROW, H. Reactions to the cessation of stimuli and their nervous mechanism. *Psychol. Rev.*, 1915, 22, 423–452.

161. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

162. WRIGHT, G. W. The latency of sensations of warmth due to radiation. *J. Physiol.*, 1951, 112, 344–358.

163. WUNDT, W. *Grundzüge der physiologischen Psychologie*. (5th Ed.) Leipzig: Engelmann, 1903.

REPRESENTATIVE vs. SYSTEMATIC DESIGN IN CLINICAL PSYCHOLOGY

KENNETH R. HAMMOND

University of Colorado

The purpose of this article is to illustrate how the application of traditional, *systematic*, rather than *representative*, experimental design to problems in clinical psychology results in unjustified conclusions.[1] Five examples, all drawn from attempts to discover the effect of the examiner on the subjects' responses, are presented. In each case it will be shown that justified conclusions regarding the problems being investigated could have been drawn if representative rather than systematic design had been used. The presentation is intended to be illustrative rather than exhaustive.

Consider first the simplest form of systematic design—the classical one-variable design. Both Brunswik (3, p. 8) and Fisher (11, p. 88) point out that this design is inherited from classical physics and both agree that revision of this procedure is necessary. Fisher, for example, remarks:

In expositions of the scientific use of experimentation it is frequent to find an excessive stress laid on the importance of varying the essential conditions only one at a time. . . . This ideal doctrine seems to be more nearly related to expositions of elementary physical theory than to laboratory practice in any branch of research. In experiments merely designed to illustrate or demonstrate simple laws, connecting cause and effect, the relationships of which with the laws relating to other causes are already known, it provides a means by which the student may apprehend the relationship, with which he is to familiarize himself, in as simple a manner as possible. By contrast, in the state of knowledge or ignorance in which genuine research, intended to advance knowledge, has to be carried on, this simple formula is not very helpful (11, p. 88).

Brunswik and Fisher agree concerning the disadvantages of classical experimental design, but they differ as to the remedy. Their differences lie not so much in the general nature of the reform needed as in the extent of the reform. Thus, although Fisher's dissatisfaction led him to develop the multivariate analysis of variance and related techniques, Brunswik's efforts have resulted in a more thorough and more radical revision of experimental methodology. For, although Fisher urged multivariation of conditions in the experiment (wherein results are obtained), thereby approaching *one* of the conditions permitting the application of results, Brunswik urges that in order to eliminate the traditional artificiality in the choice and manner of variation of the experimental variables, the conditions of the experiment *represent statistically* the universe of situations toward which one wishes to generalize.

---

[1] The two types of design are juxtaposed here in conformity with a distinction established by Brunswik (3). In the sense in which he uses the terms, "representative design" refers to the transfer of the principles of sampling statistics from the subjects of a psychological investigation to the objects or situations which constitute the stimuli in the investigation. The arbitrary orderliness with which these external (independent) variables are customarily handled is summarized by Brunswik under the opposite heading of "systematic design," and Fisher's factorial design (11) is presented as a relatively recent and relatively complex example. Fisher has also used the term "systematic" in a similar, albeit somewhat casual, manner (12).

In suggesting that the logic of sampling theory (which psychologists have long applied to populations of subjects) be applied to stimulus situations, Brunswik brings experimental methodology in line with the modern statistical approach to the problem of inductive inference. For example, "it is clear that the statistical significance of a result may be investigated in both directions" (3, p. 36). Situational (or "ecological") generality of results, therefore, is as much a challenge to statistical scrutiny as the subject-populational generality of results. As Brunswik points out, not only should we be concerned with the number of stimulus variables included in the experiment, but we must also consider the *manner of variation of the stimulus variables and their covariation among one another*, a problem which Fisher, in his concern with the mathematical problems involved in increasing the number of variables, did not consider as fully as psychologists might wish.

The above is an extremely cursory description of the fundamental notion of representative design. The examples from clinical psychology presented later, however, will further clarify its meaning. For a complete exposition of representative design, the reader is referred to Brunswik's monograph (3). From the field of academic psychology—notably from the perceptual constancies, depth perception, gestalt problems and illusions, and social perception —the reader may find concrete examples of representatively designed experiments and ecological surveys in (2; 3, pp. 24–38, 41–52; 6; 9). The general basis for representative design as given by the probabilistic nature of behavioral adjustment is presented by Brunswik in (1). Further discussion of representative design and of

its place in the historical development of psychology may be found in (5).

An analogy between the discovery, on the part of relativity physics, of the confinement of traditional physical laws to a limited universe of conditions, on the one hand, and the considerations in psychology that have led to the establishment of representative design, on the other, has been pointed out by Hammond (16). In a note to this analogy, Brunswik (4) has endorsed the writer's interpretation that in spite of the apparent stress on "theory," in the first case, and on "design," in the second, the major problem in both cases is the same, that is, "generalization."

Another paper by Hammond (15) constitutes an application of the principles of representative design to certain generalization problems frequently mismanaged in current social and clinical psychology. This paper endeavored to criticize a concrete example of a research project, an interview study by Robinson and Rohde (21). In this study the subjects had been properly sampled but the sampling of the interviewers had been ignored; yet the conclusions concerning the "significance" of the results were extended to both the subject and the interviewer populations. The point made by the present writer in criticism of this report was that the interviewers should have been treated as "objects," that is, as parts of the "situations," and thus under the rules of representative design should have been sampled in the same manner as were the subjects proper; or else the generalizations should have been limited to the subject population.[3]

[3] In pointing out the wide discrepancies between the large numbers of "judges" (responders or subjects proper) and the small numbers of "social objects" (subjects func-

In the present paper we wish to extend our criticism of research reported in the literature to the inappropriate application of systematic design to the type of experiment which seeks to determine the effect of the examiner on the subjects' responses.

ONE-FACTOR SYSTEMATIC DESIGN

In the examples discussed below, the attempt is made to vary one factor, such as the race or sex of the examiner, in order to find the effect of the stimulus variable in question on the subjects' responses.

Effect of race. Reiss, Schwartz, and Cottingham (20) were concerned mainly with verifying the utility of the (Thompson) Negro version of the TAT, but were also concerned with discovering the effect of the race of the examiner on the responses of the subjects to the Thompson TAT. "Fifteen Negro and 15 white students were tested by a Negro administrator and 15 Negro and 15 white students were tested by a white administrator" (20, p. 704). Thus, the independent variable of race is constituted by *one* representative from each race. In light of that fact, consider the following conclusion drawn by the authors: "Northern white *Ss* produce longer stories than do Northern Negroes when the stimulus material offers Negro figures *regardless of the color of the examiner*" (italics ours) (20, p. 708).

Note that this conclusion follows from the comparison of the stories of white and Negro subjects to *one* white

tioning as stimuli) typically found in the literature on social perception from photographs, Brunswik (3, p. 38 and Table 2) has suggested the use of the lower case letter *n* for the size of the responder sample and of the capital letter *N* for the size of the ecological or object sample. This system has been used in the present paper also.

and *one* Negro examiner. It is apparent that while there is a potentially adequate sample of $n = 60$ subjects, the size of the object (or ecological) sample is thus a bare $N = 2$ (for the use of *n* vs. *N* see footnote 2). If, in fact, "it is clear that the statistical significance of a result may be investigated in both directions" (3, p. 36), then we may legitimately ask how it is possible to generalize from results obtained with one representative of each race as a stimulus, any more than if the experiment included only one member from each race as a subject? Generalization from sample to population is necessary on both sides of this experiment and the statistical rules which permit generalization hold under both conditions. For Reiss *et al.* to compare the results obtained under the conditions described above with Thompson's results is just as meaningless as if the two experiments were each carried out with a single subject. Clearly, an invalid generalization is drawn in the above experiment. Only a representative type of experimental design in which adequate consideration of the situation to which the experimenter wishes to generalize is given would permit valid generalization concerning the effect of the stimuli.

Effect of sex. Curtis and Wolf report an experiment in which the problem is stated as follows: "To study the influence of the sex of the examiner on the production of sex responses on the Rorschach" (8, p. 345). The subject-population sample consisted of 586 Rorschach records. The independent variable, sex of the examiner, was constituted by three female and seven male examiners. The conclusions are: "There is a significant difference between our male and female examiners on the number of records with sex responses."

This experiment suffers from the same lack of attention to the establishment of an independent variable as the experiment discussed earlier. Note the elaborate concern of the authors to establish subject-populational generality ($n = 586$). Contrast this sampling procedure with that on the stimulus side (male $N = 7$, female $N = 3$). Yet it is the latter sample which constitutes the independent variable and which therefore leads to the conclusion that the sex of the examiner influences the response of the subject. This type of imbalance (favoring n over N) is characteristic of the indiscriminate application of systematic design. Again, there is a clearly invalid generalization on the stimulus side of the experiment which only representative design can remedy.

An approximation to representative design. An example of an "effect of the examiner" experiment which approximates representative design may be found in Gibby (14). Twelve examiners each tested 20 patients under one set of conditions, and in the second set of conditions "one hundred thirty-five subjects were used, each randomly assigned to one of nine examiners" (14, p. 450). Although this experiment also shows marked imbalance (subject n's equaling 240 and 135, and object N's equaling 12 and 9), a better approximation is obtained here than in any other experiment of this sort seen by the writer. Gibby's experiment is cited mainly to illustrate that representative design does not present insuperable difficulties.

One further point. If the experimenter wishes to limit his conclusions to the particular conditions of the experiment (as is frequently the case in applied problems), we have no criticism. If, for example, Reiss *et al.*

wish to point out that *this* examiner produces (or does not produce) different responses than *that* examiner, we have no criticism. Generalizations such as those made in the above experiments, however, demand randomization—and randomization is what is lacking.

In summary, unwarranted conclusions concerning the effect of the examiner occur as a result of failure to establish the crucial independent variable. It is important to note that failure to establish the independent variable occurs as a result of applying the logic of statistical inference to one side of the experiment only.

MULTIFACTOR SYSTEMATIC DESIGN

This section is concerned with multifactor systematic design (factorial design) methods and representative design. We begin with an example from the same field of research as in the previous section.

Sex and method of administration. Garfield, Blek, and Melker (13) were concerned with ascertaining the effect of the sex of the examiner and two methods of administration (complete session and interrupted session) upon TAT stories. Two male and two female examiners constituted the sex variable and one of each was assigned to the different methods. The experimenters also wished to discover whether differential effects would be obtained with male ($n = 54$) and female ($n = 56$) subjects. Although the experimenters do not treat their results completely in terms of a factorial design, most readers will recognize that the conditions of the experiment allow use of a $2 \times 2 \times 2$ design as in Table 1. For purposes of illustration we will assume that Garfield *et al.* did set their experiment up in this fashion. This assumption will in no way invalidate our analysis.

TABLE 1

EXAMPLE OF FACTORIAL DESIGN

Split Session				Complete Session			
Male *E*		Female *E*		Male *E*		Female *E*	
Male *Ss*	Fe *Ss*	Male *Ss*	Fe *Ss*	Male *Ss*	Fe *Ss*	Male *Ss*	Fe *Ss*

A useful, although uncustomary, step in analyzing the independent variables of experiments designed in this fashion is to separate the physical condition variables from the "person condition" variables. If, for example, the independent variables include 2 methods, 2 types of examiners, and 2 types of subjects, as does the Garfield experiment, it is important to note that we have a set of physical conditions (methods) and two sets of "person" conditions. In the experiment under discussion, the "person" conditions can be further broken down into subjects and "objects," i.e., examiners.

Person conditions (subjects). The customary principles of subject-population sampling apply here and need no discussion.

Person condition (objects). Again it is easy to see that the same restrictions concerning generalization apply here as in subject sampling. Failure to observe these restrictions results in Garfield's unjustified generalization that sex of the examiner does not influence the subjects' response. Factorial design per se does not remove these restrictions. As Edwards says in discussing a similar hypothetical factorial design experiment wherein instructors and methods constitute the independent variables, "let us suppose that we have selected the instructors to represent particular types or personalities or abilities. The three used in the experiment are definitely not a random sample from

any defined population" (**10**, p. 249). In other words, generalization to other instructors is prohibited; conclusions are confined to the *particular* instructors involved in the experiment.

Here we would like to point out that in "effect of the examiner" experiments it ordinarily will be more important to achieve preciseness with regard to the independent variable than the dependent variable. That is, we will be primarily concerned with the problem of whether or not certain examiner variables make a difference—to whom (that is, to which subject-population) they make a difference will ordinarily be of less concern. Therefore sampling of objects (examiners) will require a larger sample, more carefully considered in terms of representativeness, than will sampling of subjects.

Physical conditions. Here the issue of sampling and generalization becomes somewhat obscure. It is not easy to conceive of methods, or tests, or other physical conditions being selected by means of random sampling procedures. Yet this issue of random selection of conditions has a very important and practical bearing on factorial methods, for it can be reduced to the question of what one considers to be the error term in the design (within-groups variance or interaction variance) against which to test the significance of main effects or interaction effects. We turn now to this question.

Within-Groups Variance as Error

Under the circumstance where the physical conditions of the experiment are selected by nonsampling methods, generalization must be confined to the population of subjects sampled with respect to the *particular* physical conditions present in the experiment. Within-groups variance is legitimate as error, since here is where random sampling took place. Therefore, generalization takes place with regard to subjects only.

But—what about generalizations concerning the conditions of the experiment? Here the crucial issue between systematic factorial design and representative design lies in the *manner of covariation of the physical condition variables among one another*. Unless their *arrangement* (not merely their number) in the experiment is considered in light of the conditions toward which the generalization is aimed, our conclusions are restricted from that generalization.

Fisher was very conscious of the necessity for achieving generalization concerning the range of conditions. For example: "The exact standardization of experimental conditions, which is often thoughtlessly advocated as a panacea, always carries with it the real disadvantage that a highly standardized experiment supplies direct information only in respect of the narrow range of conditions achieved by standardization. Standardization, therefore, weakens rather than strengthens our ground for inferring a like result, when, as is invariably the case in practice, these conditions are somewhat varied" (11, p. 97). Fisher also considered the problem of the arrangement of conditions to be extremely important in multivariate designs and wrote two papers (12, Ch. 17, 28) contrasting the systematic arrangement of field conditions to random arrangement—to the disadvantage of systematic arrangement. Brunswik draws the issue more sharply, pointing out that "generalizability of results concerning . . . the variables involved must remain limited unless at least the range, but better also the distribution . . . of each variable, has been made representative of a carefully defined universe of conditions" (3, p. 53). Brunswik goes beyond Fisher in asserting that systematic factorial designs frequently "tie," or link together, physical condition variables according to the convenience (sometimes arithmetic) of the experimenter's circumstance (3, p. 6). Likewise, variables may be "untied"; that is, the links or correlations among variables in the situation to which the results are to be applied are often disrupted, usually through the "hold all other variables constant," or "isolate a variable" procedure. Note the manner in which certain stimulus variables are arbitrarily "tied" and "untied" in the following example.

Personality and sex. Holtzman (17) was interested in ascertaining the effect of the examiner as a variable in the Draw-A-Person Test. His procedure in establishing the independent variables of sex and of personality was as follows:

Four experienced examiners, two male and two female, were chosen from a group of advanced graduate students in clinical psychology. The two pairs of examiners were selected so as to maximize differences in examiner appearance and personality within both sexes. Examiner M1 was nearly a foot taller and sixty pounds heavier than the other male examiner, M2. The two female examiners, F1 and F2, were approximately the same size but differed considerably in feminine qualities" (17, p. 145).

In line with our general emphasis

on the need for scrutinizing the manner in which the independent variable is constituted in these experiments, note that height and weight were "separated" or varied "within sex" for the male examiners (M1 was a foot taller and sixty pounds heavier than M2). On the other hand, the height of the female examiners was the same (their weights are not mentioned), but their "feminine qualities" "differed considerably." Apparently several different stimulus variables are "tied" (height with female sex) and "untied" (height and weight from male sex and "feminine qualities" within female sex). This situation obscures the independent variable. This obscurity is bound to result when randomization is not effected in the sampling procedures, whether we are dealing with sampling of subjects or "objects" (in this case, examiners).[3]

Thus, when conditions are not samples, generalization is limited to the subject-population and therefore within-group variance is the only legitimate error term.

INTERACTION TERMS AS ERROR

Now consider the circumstance

[3] Note the hypothesis to be tested under these circumstances "(1) The sex of the examiner has a measurable effect ... (2) The personal characteristics of the examiner aside from sex have a measurable effect ..." (17, p. 145). The "male variable" is established by drawing two males from the population, one each from some hypothetical personality-type population. The same for the "female variable." From this sample of examiners Holtzman attempts to refute the findings of Sinnett and Eglash (22) (who also used two examiners) concerning the relationship between examiner personality and response to the Draw-A-Person Test. Obviously these particular examiners differed in many variables other than sex characteristics. Such variables can be eliminated as "causes" of results only by randomization—exactly as they are eliminated in sampling subject populations.

where the interaction term might be considered as error. This is the point at which textbooks begin to meet the problem of random sampling of physical conditions. In his discussion of the interaction term as error in his *Design of Experiments* (11, pp. 201-205), Fisher deals with the problem of random sampling of conditions almost to the exclusion of other topics. Both Edwards (10) and Lindquist (18) consider this problem in connection with psychological and educational experiments in which methods, schools, etc. are conditions of the experiment.

Edwards (10, p. 252) in discussing interactions as error terms states, " ... it would be illogical to argue that the ... particular methods ... selected for investigation have been randomly selected from a population of methods." Lindquist (18, p. 169) discusses this problem through a hypothetical experiment where style and size of type are the independent variables. " ... the particular *styles* (or *sizes*) involved may not strictly be considered as a random sample from a 'population' of styles (or sizes). The interaction variance in a factorial design is therefore usually not strictly a measure of normally distributed *random* fluctuations, which theoretically must be true of the *error* term in any F-test or t-test."

When discussing the necessity for randomness in the use of interaction terms as error terms, Edwards makes very explicit the fact that it is "illogical to argue" that particular methods have been "randomly selected" and that particular instructors cannot be considered random samples. Yet he chooses as an example an experiment by Child (7) which violates this requirement (10, p. 261). Edwards describes the conditions as follows: "The variables introduced were as follows: the sex of the children used

...subjects in the experiment; the sex of the experimenter present during the test situation; the nature of the barrier introduced between the subject and the distant goal object; and the type of instructions given to the child." Child recognizes that his conclusion that "choice of a distant goal was more frequent in the presence of a woman experimenter than in the presence of a man experimenter" is "severely limited by the fact that the sample of experimenters was limited to one of each sex" (7, p. 30). However, Child does use the pooled interaction term as error. "The 11 degrees of freedom pertaining to the 11 possible interactions were therefore pooled for an estimate of error" (7, p. 19). If the writer has correctly identified Child's interaction terms, four ($E\times$barrier$\times$instructions, $E\times$barrier, $E\times$instructions, barrier$\times$instructions) of the 11 interactions do not include a variable selected at random. As Edwards put it, "furthermore, and this is most important, it is necessary that the categories of one or more of the variables in the experimental design be a random selection from the population being sampled" (10, p. 252). This is followed by a footnote: "This condition will not be met by argument after the experiment has been carried through to completion."

In summary, then, random sampling of the conditions, physical or otherwise, of the experiment is prerequisite to the use of interaction terms as error variance. Ignoring this prerequisite violates the principle of generalization through randomization in exactly the same fashion as in the case of a nonsampling one-factor design.

AGRICULTURAL DESIGNS IN PSYCHOLOGY

The above remarks concerning interaction terms lead directly to another issue—the use of agricultural designs in psychology. For example, in Lindquist's discussion of interaction terms as error, after very clearly pointing out the logic involved in compromising with the statistical assumptions involved, he concludes the discussion by emphasizing that " . . . the procedure just recommended is arbitrary in character, *although wide experience in agricultural research indicates that it is usually satisfactory*" (italics ours) (18, p. 170).[4] Snedecor, however, while discussing interaction terms as error terms in connection with an agricultural experiment, says, "The requirement of randomness is apparently never met in this kind of work, so that statements about probability must be considered inexact" (23, p. 303).

The above remarks concerning agricultural research have been emphasized because they are the key to a fundamental issue. The issue is whether the application of agricultural experimental procedures to psychological experiment needs further scrutiny. We believe that such scrutiny is in order on grounds that there are important discrepancies between the aims and conditions of most agricultural research and most psychological research.[5] The following points bear on this question.

The type of agricultural research for which Fisher developed factorial design was primarily *engineering-type* research. He was concerned " . . . that, in any case, there will be no reason for rejecting the experi-

[4] Lindquist also notes (18, p. 169) that there will be some circumstances where this procedure will not be reasonable.

[5] See McNemar (19) for a discussion of an important difference in the conditions of agricultural and psychological research in connection with latin-square design.

mental results on the ground that the test was made in conditions differing in one or other of these respects from those in which it is proposed to apply the results" (11, p. 98). In engineering-type experiments the selection and arrangement of the conditions are usually dictated by the particular circumstances to which the experimenter wishes to *apply* his results—and to which he confines his results and conclusions, as Fisher made clear. A further characteristic of engineering-type experiments is that *future* conditions are under the control of the experimenter, so that he can maintain the conditions of the experiment, e.g., issuing the same ratios of ingredients in the ration, the same ratios of chemicals in the fertilizer, etc., as used in the experiment. Therefore, Fisher's factorial design methods are eminently appropriate to the applied problems with which he was concerned. Where the problem of generalizing beyond a particular plot—the soil heterogeneity problem—was concerned, however, he was definite and explicit about the advantages of randomization. (See 11, 12.)

Psychologists, however, have other problems at stake than those of application; for example, it is in the nature of the theoretical task to seek generality. It seems legitimate, therefore, to ask if psychologists really wish to confine their conclusions to the particular conditions of the experiment, in which the manner or arrangement of the variables is frequently due to convenience, or as Fisher put it, due to the "thoughtless advocation of standardization as a panacea" (11, p. 97). Moreover, if the future conditions to which the psychologist wishes to predict are not under his control, it also seems legitimate to ask if the manner of covariation of the variables has been maintained in the experiment as in the future situation. If the psychologist does not wish to accept the limitations of systematic design, however, a thorough scrutiny of the principles of representative design is in order, for it is precisely the question of generalization with which representative design is concerned.

SUMMARY

This paper illustrates how unwarranted conclusions may be reached through the application of traditional systematic design to a given problem in clinical psychology—the effect of the examiner on the subjects' responses. Both one-factor and multifactor designs are discussed.

REFERENCES

1. BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, 50, 255–272.
2. BRUNSWIK, E. Distal focussing of perception. Size-constancy in a representative sample of situations. *Psychol. Monogr.*, 1944, 56, No. 1 (Whole No. 254).
3. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: Univer. of California Press, 1947. (Also in J. Neyman [Ed.], *Berkeley symposium on mathematical statistics and probability*. Berkeley: Univer. of California Press, 1949. Pp. 143–202.)
4. BRUNSWIK, E. Note on Hammond's analogy between "relativity and representativeness." *Phil. Sci.*, 1951, 18, 212–217.
5. BRUNSWIK, E. The conceptual framework of psychology. Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.*, v. 1, No. 10.)
6. BRUNSWIK, E., & KAMIYA, J. Ecological cue validity of "proximity" and of other Gestalt factors. *Amer. J. Psychol.*, 1953, 66, 20–32.

7. CHILD, I. L. Children's preference for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, **60**, No. 4 (Whole No. 280).

8. CURTIS, H. S., & WOLF, E. B. The influence of the sex of the examiner on the production of sex responses on the Rorschach. *Amer. Psychologist*, 1951, **6**, 345. (Abstract)

9. DUKES, W. F. Ecological representativeness in studying perceptual size-constancy in childhood. *Amer. J. Psychol.*, 1951, **64**, 87–93.

10. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

11. FISHER, R. A. *The design of experiments.* (4th Ed.) New York: Hafner, 1947.

12. FISHER, R. A. *Contributions to mathematical statistics.* New York: Wiley, 1950.

13. GARFIELD, S. L., BLEK, L., & MELKER, F. The influence of method of administration and sex differences on selected aspects of TAT stories. *J. consult. Psychol.*, 1952, **16**, 140–144.

14. GIBBY, R. G. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.*, 1952, **16**, 449–455.

15. HAMMOND, K. R. Subject and object sampling—a note. *Psychol. Bull.*, 1948, **45**, 530–533.

16. HAMMOND, K. R. Relativity and representativeness. *Phil. Sci.*, 1951, **18**, 208–211.

17. HOLTZMAN, W. H. The examiner as a variable in the Draw-A-Person Test. *J. consult. Psychol.*, 1952, **16**, 145–148.

18. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.

19. McNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.

20. REISS, B. F., SCHWARTZ, E. K., & COTTINGHAM, ALICE. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700–709.

21. ROBINSON, D., & ROHDE, S. Two experiments with an anti-Semitism poll. *J. abnorm. soc. Psychol.*, 1946, **41**, 136–144.

22. SINNETT, E. R., & EGLASH, A. The examiner-subject relationships as a variable in the Draw-A-Person Test. Paper read at the Midwest. Psychol. Ass., Detroit, May, 1950.

23. SNEDECOR, G. W. *Statistical methods.* (4th Ed.) Ames, Iowa: State Coll. Press, 1946.

KOLMOGOROV-SMIRNOV TESTS FOR PSYCHO-
LOGICAL RESEARCH

LEO A. GOODMAN[1]
University of Chicago

In an excellent paper, Moses (13) presents some of the principal non-parametric methods and an intuitive explanation of their rationale, properties, and applicability, with a view to facilitating their use by workers in psychological research. One of the important topics in the field of non-parametric methods is the Kolmogorov-Smirnov statistic. Recent results and tables on this topic have been prepared which contribute toward establishing the Kolmogorov-Smirnov statistic as a standard nonparametric tool of statistical analysis.

We shall present an intuitive explanation of the rationale and uses of the Kolmogorov-Smirnov statistic. A table will be presented which facilitates the use of the Kolmogorov-Smirnov statistic by research workers. Some illustrative examples will also be given.

1. Two-Sided Tests

1.1. *One-Sample Test*

In this section tests of "goodness of fit" will be considered. That is, we shall be concerned with the agreement between the distribution of a set of sample values and a theoretical distribution. Probably the most widely used nonparametric test of "goodness of fit" is the chi-square test (4). However, some evidence has been presented indicating that the test which we shall now describe, the

Kolmogorov-Smirnov test for goodness of fit, may be a better all-around test, when it is applicable, than the chi-square test (1, 12). A concise description of other tests of fit appears in (2).

Suppose that a population is thought to have some completely specified cumulative frequency distribution function, say $F(x)$. That is, for any specified value of x, the value of $F(x)$ is the proportion of individuals in the population having measurements less than or equal to x. The observed cumulative step-function $S_N(x)$ of a sample of N observations (that is, $S_N(x) = k/N$, where k is the number of observations less than or equal to x) is expected to be fairly close to this completely specified distribution function $F(x)$. If it is not close enough, we have evidence that the hypothetical distribution $F(x)$ is not the correct one. As a measure of how far the observed cumulative step-function $S_N(x)$ is from the hypothetical distribution $F(x)$ we use the maximum absolute difference d between $S_N(x)$ and $F(x)$; that is, $d = \text{maximum}_x |F(x) - S_N(x)|$. When d is large we have evidence that $F(x)$ is not the correct population cumulative frequency distribution function.

Let us first consider the case where $F(x)$ is a continuous cumulative distribution function. If the correct population cumulative distribution is in fact $F(x)$, then the sampling distribution of d is known and has been tabled (1). For example, if $F(x)$ is the correct population distribution, we find from p. 428 in (1) that the prob-

ability that $d \geq 5/15$ is $Pr\{d \geq 5/15\}$ $= 1 - Pr\{d < 5/15\} = 1 - .945 = .055$, and $Pr\{d \geq 6/15\} = 1 - .989 = .011$ for a sample of $N = 15$ observations.

Hence, in order to test the null hypothesis that $F(x)$ is the correct population distribution using a sample of $N = 15$ observations, the maximum absolute deviation d between the sample cumulative step-function $S_{15}(x)$ and $F(x)$ is computed. The null hypothesis is rejected when $d \geq 5/15$, if the null hypothesis is tested at the .055 level of significance. If the test is at the .011 level of significance, the null hypothesis is rejected when $d \geq 6/15$.

Let us now consider the case where $F(x)$ may be a discontinuous cumulative distribution function. Several articles dealing with the Kolmogorov-Smirnov tests claim that the methods apply only where the chance variable is continuous. This is true only if exact probability statements are required. (The reader will note that exact probability statements cannot be made even if the chi-square statistic is used to test goodness of fit since little is known about the actual sampling distribution of the chi-square statistic for finite sample size N and given $F(x)$. The chi-square statistic becomes approximately distribution-free when the sample size N approaches infinity, but is not distribution-free for finite N.) The inequalities stated in (5, 7) serve to validate the use of the Kolmogorov-Smirnov statistic in the case where $F(x)$ may be discontinuous. From these inequalities we see that when $F(x)$ may be discontinuous, the error obtained will be in the "safe direction" if tables are used which assume that $F(x)$ is continuous. More precisely, the level of significance of a test based on the sampling distribution of d will be no larger than the level of significance of that test when

$F(x)$ is assumed continuous. For example, if $F(x)$, which may be discontinuous, is the correct population distribution function, we find from the tables on p. 428 of (1) that the probability that $d \geq 5/15$ is at most .055, and $Pr\{d \geq 6 | 15\} \leq .011$ for a sample of $N = 15$ observations. Hence, in order to test the null hypothesis that $F(x)$, which may be discontinuous, is the correct population distribution, the value of d is computed. If the null hypothesis is rejected when $d \geq 5/15$, then the level of significance is, at most, .055. The test will be at no more than the .011 level of significance if the null hypothesis is rejected when $d \geq 6/15$. Hence, if the same test is used as when $F(x)$ was assumed continuous, the level of significance will be no more than .055 (or .011).

The Kolmogorov-Smirnov statistic can also be used to estimate probabilities and obtain confidence bands for the true cumulative distribution function $F(x)$ (1, 7, 10, 15, 16). These confidence bands will be free from any restriction concerning the nature of the function $F(x)$. We need not assume that $F(x)$ is continuous in order to obtain confidence bands for $F(x)$. Let us illustrate this use with a particular set of data. A sample of 15 observations was obtained. If this sample is arranged in order of increasing size, we obtain

1, 2, 2, 2, 2, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5.

The reader may be interested to know how these 15 observations were obtained. A student was asked to list three men whom he liked and three men whom he did not like from among all the men he had known since birth. The number 0 was assigned to the person liked the best, number 1 to the person liked second best, number 2 to the person liked third best, and so on, and 5 was as-

signed to the person disliked the most. He was then asked to place the number 0 on a sheet of paper if he thought that his number 0 person (best-liked person) would be the richest (among the total of six men listed). He was to place the number 1 on the paper if he thought that his number 1 person (second best liked) would be richest, and so on, and the number 5 if he thought that the person he disliked most would be the richest among the six. Hence, a single number from 0 to 5 was obtained from this student. Fifteen such observations were obtained by using fifteen different students as the subjects. The sample cumulative step-function $S_{15}(x)$ for the 15 observations is given in Fig. 1. From the tables on p. 428 in (1) we see that the chance is at least .945 that the maximum deviation between $S_{15}(x)$ and the true cumulative distribution $F(x)$
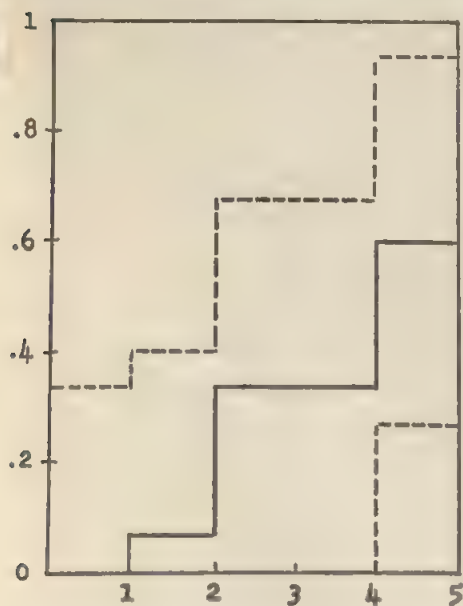
FIG. 1. TWO-SIDED 94.5 PER CENT CONFIDENCE BAND FOR $F(x)$ OBTAINED FROM A SAMPLE OF 15 OBSERVATIONS

will be less than 5/15. Hence, if a band of width 5/15 is drawn above and below $S_{15}(x)$, we can state with at least "94.5 per cent confidence" that the true cumulative distribution $F(x)$ lies within that band. The 94.5 per cent confidence band for $F(x)$ is illustrated in Fig. 1 by the dotted lines above and below $S_{15}(x)$. Therefore, any number of statements of the following kind may be made simultaneously with at least 94.5 per cent confidence: $F(0) < 1/3$, $F(3) < 2/3$, and $F(4)$ is a number between 26.667 per cent and 93.333 per cent.

Let us consider the null hypothesis that a student is equally likely to choose any one of the six numbers. Then the population cumulative distribution function $F(x)$ would be as given in Table 1. The values of $S_{15}(x)$ and $|F(x) - S_{15}(x)|$ are also given in the table. Since the maximum absolute difference between $F(x)$ and $S_{15}(x)$ is $10/30 = 5/15$, the null hypothesis is rejected at the .055 level of significance.

The reader will note that the significance test which was performed was for a completely specified population cumulative distribution function $F(x)$. In cases where parameters must be estimated from the sample (for example, when the null hypothesis is that the population distribution is normal with unspecified mean and standard deviation, and the mean and standard deviation must first be estimated from the sample), there are no theoretical results at present which give exact critical levels for the Kolmogorov-Smirnov statistic. The distribution of d is not known when certain parameters of the population have been estimated from the sample. It may be expected, however, that the effect of adjusting the population mean and standard deviation to those of the sample will be to reduce the

TABLE 1

ABSOLUTE DIFFERENCE BETWEEN A SPECIFIED $F(x)$ AND $S_{15}(x)$ FOR A SAMPLE OF 15 OBSERVATIONS

x	0	1	2	3	4	5
$F(x)$	1/6	2/6	3/6	4/6	5/6	1
$S_{15}(x)$	0	1/15	5/15	5/15	9/15	1
$\lvert F(x)-S_{15}(x)\rvert$	5/30	8/30	5/30	10/30	7/30	0

critical level of d. If the critical value of d (from tables which assume a completely specified population distribution) is exceeded in these circumstances, we may safely conclude that the discrepancy is significant (see p. 73 of [12]). In cases where parameters must be estimated from the sample, the chi-square test is easily modified by reducing the number of degrees of freedom. The Kolmogorov-Smirnov test has no such known modifications.

1.2. *Two-Sample Test*

In this section the problem of testing whether two random samples have been drawn from the same population is considered. That is, we shall be concerned with the agreement between the distributions of two sets of sample values.

Let us denote the observed cumulative step-function of the first sample of N observations by $S_N(x)$, and let $S'_M(x)$ be the observed cumulative step-function of the second sample of M observations. The two cumulative step-functions $S_N(x)$ and $S'_M(x)$ are expected to be fairly close to each other if both samples are drawn from the same population. If they are not close enough, we would have evidence that the samples come from different populations. (That is, the population cumulative distribution function for the values from the first sample is different from the population cumulative

distribution for the values from the second sample.) As a measure of how far apart are the two cumulative step-functions we use the maximum absolute difference d' between them; that is, $d' = \text{maximum}_x \ \lvert S_N(x) - S'_M(x)\rvert$. When d' is large we have evidence that the samples came from different populations.

Let us first consider the case where the values from both samples are assumed to have continuous population cumulative distributions $F(x)$ and $G(x)$, respectively. We wish to test the null hypothesis that $F(x) \equiv G(x)$ and the null hypothesis will be rejected if the observed value of d' is significantly large. The limiting distribution of d' has been tabled in (15), and a method of obtaining the exact distribution of d' for small samples has been given (11) when in fact $F(x) \equiv G(x)$. A short table for equal size samples is also available (11). The explicit expression for the distribution function of d' has been given recently in (6) for equal size samples. From the tables on p. 126 of (11) we find that, say, in the case where $M = N = 15$, the probability that $d' \geq 7/15$, when in fact $F(x) \equiv G(x)$, is $Pr\{d' \geq 7/15\} = 1 - Pr\{d' \leq 6/15\} = 1 - .925 = .075$, and $Pr\{d' \geq 8/15\} = 1 - .974 = .026$. Hence, in order to test the null hypothesis that $F(x) \equiv G(x)$ at the .075 level of significance, the value of d' is computed and the null hypothesis is rejected when $d' \geq 7/15$. If the test is at the

.026 level of significance, the null hypothesis is rejected when $d' > 8 \cdot 15$.

Let us now consider the case where $F(x)$ and $G(x)$ may be discontinuous cumulative distribution functions. The inequalities stated in (7) serve to validate the use of the Kolmogorov-Smirnov statistic in the case where $F(x)$ and $G(x)$ may be discontinuous. From these inequalities, we see that when $F(x)$ and $G(x)$ may be discontinuous, the error obtained will be in the "safe direction" if tables are used (11) which assume that $F(x)$ and $G(x)$ are continuous. For example, suppose two samples are drawn each containing 15 observations ($M = N = 15$). In order to test the null hypothesis that $F(x) \equiv G(x)$, the value of d is computed. If the null hypothesis is rejected when $d' \geq 7/15$, the level of significance will be, at most, .075. The test will be at no more than the .026 level of significance if the null hypothesis is rejected when $d' \geq 8/15$. The tests will be free from any restriction concerning the nature of the functions $F(x)$ and $G(x)$. We need not assume that the functions $F(x)$ and $G(x)$ are continuous in order to obtain tests of the hypothesis that $F(x) \equiv G(x)$.

Let us illustrate the problem of testing whether two samples have been drawn from the same population. We shall study the agreement between the sample of 15 observations which was described in the preceding section (numbers from 0 to 5 obtained by using fifteen students as

subjects of an inquiry) and a second sample of 15 observations. If the second sample is arranged in order of increasing size, we obtain

0, 0, 0, 0, 1, 1, 2, 2, 2, 2, 3, 3, 5, 5, 5

(The reader may be interested to know that these 15 observations were obtained by using fifteen businessmen as the subjects of the inquiry. That is, each businessman was interrogated in the same manner as the students. The businessman was then asked to choose the person who would be the richest among the six men listed in the order of his preference.) The values of the cumulative step-functions $S_{15}(x)$ and $S'_{15}(x)$ for the first and second samples respectively are given in Table 2. The values of $|S_{15}(x) - S'_{15}(x)|$ are also given in the table. Since the maximum absolute difference between $S_{15}(x)$ and $S'_{15}(x)$ is 7/15, the null hypothesis is rejected at the .075 level of significance.

2. One-Sided Tests

2.1. One-Sample Test

In Section 1.1 the statement was made that the Kolmogorov-Smirnov test for goodness of fit may be a better all-around test, when it is applicable, than the chi-square test. By an all-around test of goodness of fit, we mean a test of the null hypothesis that the observed sample was drawn from a completely specified population without specifying the nature of

TABLE 2

ABSOLUTE DIFFERENCE BETWEEN THE CUMULATIVE STEP-FUNCTIONS FOR TWO SAMPLES EACH CONTAINING 15 OBSERVATIONS

x	0	1	2	3	4	5
$S_{15}(x)$	0	1/15	5/15	5/15	9/15	1
$S'_{15}(x)$	4/15	6/15	10/15	12/15	12/15	1
$|S_{15}(x) - S'_{15}(x)|$	4/15	5/15	5/15	7/15	3/15	0

the alternate hypotheses. That is, the null hypothesis that the observed sample was drawn from a completely specified population is tested against the alternate hypothesis that it was not drawn from that population. In some particular problems more specific alternate hypotheses may be desirable. For example, very often we want to decide not whether an experimental group is the same or different from the general population, but whether the experimental group is better than the general population, or more adjusted than the general population, etc. In such cases, the null hypothesis that there is no difference would be tested against the alternate hypothesis that the experimental group is better, or more adjusted, etc.

Let us consider the numerical illustration presented in Section 1.1 where the null hypothesis that a student is equally likely to choose any one of the six numbers 0, 1, 2, 3, 4, 5 was tested against the alternate hypothesis that the student is not equally likely to choose any one of the six numbers. For this particular problem and for the student body which was under investigation it seemed reasonable to expect that either (a) the six numbers would be equally likely or (b) there would be a tendency for the students to assign the higher numbers. Hence, in this case it is desirable to test the null hypothesis that the six numbers were equally likely against the alternate hypothesis that there was a tendency for the students to assign the higher numbers.

Let us try to make more precise the statement that "there would be a tendency for the students to assign higher numbers." If the six numbers were equally likely, then the proportion of the population assigning the number 0 would be $F(0) = 1/6$. Also

the proportion assigning the number 0 or 1 would be $F(1) = 2/6$, and the proportion assigning numbers no larger than 2 would be $F(2) = 3/6$. Similarly, $F(3) = 4/6$, $F(4) = 5/6$, and $F(5) = 1$. If "there is a tendency to assign higher numbers," then the proportion $G(0)$ of the population assigning the number 0 would be no more than $1/6$ (the case where all numbers are equally likely). Also the proportion $G(1)$ assigning the number 0 or 1 would be no more than $2/6$, and the proportion $G(2)$ assigning the numbers no larger than 2 would be no more than $3/6$. Similarly $G(3) \leq 4/6$, $G(4) \leq 5/6$, and $G(5) \leq 1$. Hence, the statement, "there would be a tendency for the student to assign higher numbers," may be replaced by the statement that the true population cumulative distribution function $G(x)$ is no more than $F(x)$; that is, $G(x) \leq F(x)$ for all values of x.

The null hypothesis that the true population cumulative distribution is, in fact, the specified $F(x)$ is to be tested against the alternate hypothesis that the true population cumulative $G(x)$ is no more than $F(x)$ (with the "less than" relation holding for some values of x). This problem may be considered the one-sided analog of the problem discussed in Section 1.1 where the null hypothesis that the true population cumulative distribution is in fact the specified $F(x)$ is tested against the alternate hypothesis that the true population cumulative is not $F(x)$. As a measure of how far the observed cumulative step-function $S_N(x)$ is from the hypothetical distribution $F(x)$ we use the maximum difference c between $S_N(x)$ and $F(x)$. That is, $c = \text{maximum}_x [F(x) - S_N(x)]$, which is the one-sided analog of the maximum absolute difference d. When c is large, we have evidence that $F(x)$ is not the correct population cumulative distribution

and that the true population cumulative $G(x)$ is no more than $F(x)$. If the correct population cumulative is continuous and is in fact $F(x)$, then the sampling distribution of c is known, and the explicit expression for the distribution function of c is given by equation 3.0 on p. 593 in (3). Using equation 3.0, we find for example that when $N = 15$ the chance that c will be no more than 5/15 is

$$P_{15}(5/15) = 1 - \frac{1}{3} \sum_{j=0}^{10} \binom{15}{j}$$
$$\left(\frac{2}{3} - \frac{j}{15}\right)^{10-j} \left(\frac{1}{3} + \frac{j}{15}\right)^{j-1} = .97.$$

Hence, the probability that $c \geq 5/15$ is $Pr\{c > 5/15\} = 1 - .97 = .03$. In order to perform a "one-sided test" of the null hypothesis that $F(x)$ is the correct population distribution using a sample of $N = 15$ observations, the value of c is computed. The null hypothesis is rejected when $c \geq 5/15$ if the test is at the .03 level of significance.

For the particular numerical illustration presented in Section 1.1, we find that $c = 5/15$. Hence, the null hypothesis that a student is equally likely to choose any one of the six numbers is rejected at the .03 level of significance, and the alternate hypothesis that there is a tendency to assign higher numbers is accepted.

In Section 1.1 a method was given for obtaining two-sided confidence bands for the true cumulative distribution function $F(x)$. This method may be modified in order to obtain one-sided confidence bands for $F(x)$. For example, Fig. 1 gives a two-sided 94.5 per cent confidence band for $F(x)$ obtained from a sample of 15 observations. We might make the one-sided confidence statement that the true cumulative distribution $F(x)$ lies below the upper limit of the band

presented in Fig. 1. From the results presented earlier in this section we see that there is at least "97 per cent confidence" that $F(x)$ lies below the upper limit of the band in Fig. 1.

2.2. *Two-Sample Test*

In this section a one-sided analog of the problem discussed in Section 1.2 will be considered. We wish to test the null hypothesis that the cumulative distribution function $F(x)$ for the values from the first sample is equal to the cumulative distribution function $G(x)$ for the values from the second sample. In other words, the null hypothesis is that $F(x) \equiv G(x)$. The alternate hypothesis to be considered is of a more specific nature than the alternate hypothesis for the all-around test presented in Section 1.2. We shall be concerned with the alternate hypothesis that $F(x) \leq G(x)$. If the alternate hypothesis is, in fact, true (with the "less than" relation holding for some values of x), we say that the population values from which the first sample was drawn are stochastically larger than the population values from which the second sample was drawn. The importance of such alternate hypotheses has been stressed in (8, 9). For example, very often we want to decide not whether the experimental group is the same or different from the control group but whether the experimental group is better than the control group, or more adjusted than the control group, etc. In such cases the null hypothesis that $F(x) \equiv G(x)$ would be tested against the alternate hypothesis that $F(x) \leq G(x)$. For this one-sided analog of the problem described in Section 1.2, we use the maximum difference c' between the observed cumulative step-functions $S_N(x)$ and $S'_M(x)$ of the first sample of N observations and of the second sample of M observations. That is, $c' = $ maximum,

$[S'_M(x) - S_N(x)]$ which is the one-sided analog of the maximum absolute difference d'. When c' is large, we have evidence that $F(x)$ is not equal to $G(x)$ and that the population values from which the first sample was drawn are stochastically larger than the population values from which the second sample was drawn. If the population cumulative functions are continuous and in fact $F(x) = G(x)$, then the limiting distribution of c' is known (9, 14). If $F(x) = G(x)$, we find that the sampling distribution of $4(c')^2 MN/(M+N)$ will have approximately a chi-square distribution with two degrees of freedom when M and N are large and M/N is not too close to either zero or infinity. Hence, the tables of the chi-square distribution may be utilized to test the null hypothesis when M and N are large. When M and N are small, the exact distribution of c' may be computed by extending the counting method presented in (11) for the two-sided problem. The explicit expression for the distribution function of c' has been given recently (6) for equal-size samples. Table 3 may be used to test the null hypothesis at either the 20 per cent, 10 per cent, 5 per cent, 1 per cent, or 0.1 per cent level of significance if $M = N$. The table gives the critical value of $c'N$ at the various levels of significance. For example, when $N = 15$ we see from Table 3 that 7 is the critical value of $c'N$ at the 5 per cent level of significance. Hence, the null hypothesis is rejected at the 5 per cent level of significance if $c' \geq 7/15$. Using the explicit expression (see [6]) for the distribution function of c', we find that

$$Pr\{c' \geq 7/15\} = 1 - Pr\{c' < 7/15\}$$

$$= \binom{30}{8} \Big/ \binom{30}{15} = .038.$$

Let us reconsider the numerical il-

TABLE 3

Critical Values of $c'N$ for the α 100 Per Cent Level of Significance

N	001	01	05	1	2
1	∞	∞	∞	∞	∞
2	∞	∞	∞	∞	2
3	∞	∞	3	3	3
4	∞	∞	4	4	3
5	∞	5	4	4	3
6	∞	6	5	4	4
7	7	6	5	5	4
8	8	6	5	5	4
9	8	7	6	5	4
10	9	7	6	5	5
11	9	8	6	6	5
12	9	8	6	6	5
13	10	8	7	6	5
14	10	8	7	6	5
15	10	9	7	6	5
16	11	9	7	7	6
17	11	9	8	7	6
18	11	10	8	7	6
19	12	10	8	7	6
20	12	10	8	7	6
21	12	10	8	7	6
22	13	11	9	8	6
23	13	11	9	8	7
24	13	11	9	8	7
25	13	11	9	8	7
26	14	11	9	8	7
27	14	12	9	8	7
28	14	12	10	9	7
29	14	12	10	9	7
30	15	12	10	9	7
35	16	13	11	9	8
40	17	14	11	10	9
45	18	15	12	11	9
50	19	16	13	11	9

lustration presented in Section 1.2. Since $c' = 7/15$, we reject the null hypothesis at the .038 level of significance that the population distribution of the numbers assigned by the students was the same as the population distribution of the numbers assigned by the businessmen and accept the alternate hypothesis that the numbers assigned by the students were stochastically larger than those assigned by the businessmen. In other words, we accept the alternate

hypothesis that there was a greater tendency for the students to assign higher numbers than the businessmen.

It is interesting to note that if approximate critical values are computed using the chi-square (with two degrees of freedom) approximation the approximate critical values are always more than the exact critical values (in Table 3) minus 1. Hence, the error in using the chi-square approximation is always in the "safe direction" for the levels of significance in Table 3 even when the sample size is small. In other words, if the null hypothesis is rejected using the chi-square approximation, it would also be rejected if exact computations had been made. It is also interesting to note that only in the following cases will the chi-square approximation lead to acceptance when an exact computation based on Table 3 would lead to rejection: $\alpha = .001$ and $N = 12$, 15, 18, 21, 25, 29; $\alpha = .01$ and $N = 8$, 14.

REFERENCES

1. BIRNBAUM, Z. W. Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample sizes. *J. Amer. statist. Ass.*, 1952, **47**, 425–441.

2. BIRNBAUM, Z. W. Distribution-free tests of fit for continuous distribution functions. *Ann. math. Statist.*, 1953, **24**, 1–8.

3. BIRNBAUM, Z. W., & TINGEY, F. H. One-sided confidence contours for probability distributions functions. *Ann. math. Statist.*, 1951, **22**, 592–596.

4. COCHRAN, W. G. The χ^2 test of goodness of fit. *Ann. math. Statist.*, 1952, **23**, 315–345.

5. DAVID, H. T. Discrete populations and the Kolmogorov-Smirnov tests. Unpublished report SRC-21103D27, Statist. Res. Cent., Univer. of Chicago.

6. GNEDENKO, B. V., & KOROLYUK, V. S. On the maximum discrepancy between two empirical distributions. *Doklady Akad. Nauk S.S.S.R. (N.S.)*, 1951, **80**, 525–528. (A review of this article appears in *Mathematical Reviews*, 1952, **13**, 570.)

7. KOLMOGOROV, A. Confidence limits for an unknown distribution function. *Ann. math. Statist.*, 1941, **12**, 461–463.

8. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

9. MARSHALL, A. W. A large-sample test of the hypothesis that one of two random variables is stochastically larger than the other. *J. Amer. statist. Ass.*, 1951, **46**, 366–374.

10. MASSEY, F. J., JR. A note on the estimation of a distribution function by confidence limits. *Ann. math. Statist.*, 1950, **21**, 116–119.

11. MASSEY, F. J., JR. The distribution of the maximum deviation between two sample cumulative step-functions. *Ann. math. Statist.*, 1951, **22**, 125–128.

12. MASSEY, F. J., JR. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. statist. Ass.*, 1951, **46**, 68–78.

13. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.

14. SMIRNOV, N. Sur les écarts de la courbe de distribution empirique. *Recueil Mathematique (Mathematiceskii Sbornik, M.S.)*, 1939, **48**, 3–26.

15. SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. *Ann. math. Statist.*, 1948, **19**, 279–281.

16. WALD, A., & WOLFOWITZ, J. Confidence limits for continuous distribution functions. *Ann. math. Statist.*, 1939, **10**, 105–118.

REMARK ON "A QUALIFICATION IN THE USE OF ANALYSIS OF VARIANCE"

VICTOR H. DENENBERG

Human Resources Research Office, Fort Knox, Kentucky

In their original article (7) Webb and Lemmon stated that the results of the over-all F test may not agree with results obtained by subsequent use of the t test applied to individual means when a functional relationship exists between the independent and dependent variables. They gave hypothetical examples which, they thought, illustrated their point. Patterson (5) and Diamond (1) have both taken exception to the original article, but Webb and Lemmon in their reply (8) stated that " . . . most of their [Patterson and Diamond's] criticisms have not been aimed directly at the core of our problem. Most of their discussion seems to deal with problems of the conventional analysis of variance situation, in which the means of the groups show no trend, but are randomly related to each other. . . . We were concerned, however, with a fairly common experimental design in which the means of the groups show a definite ordering or trend, with respect to some other experimental variable."

In all four of the articles written by the above authors, analysis of variance has been discussed as though its only use were to test for significance of differences between means. This is probably what analysis of variance has been most often used for in psychological research, but it has other uses which are just as important as the test of differences between means. One of the other uses of analysis of variance is that of testing to see whether there is a *significant linear*

or curvilinear regression between the independent and dependent variables. This is the test which should be used with Webb and Lemmon's data since they hypothesize that a functional relationship exists. This test may be performed by an extension of the conventional analysis of variance procedure. Fisher's (2) method of orthogonal polynomials is convenient to use for this analysis. Snedecor (6, Ch. 14, 15) discusses in detail the procedures necessary to make this analysis, and Johnson and Tsao (3) give a very complete discussion of the procedures involved in fitting orthogonal polynomials to a $4 \times 7 \times 2 \times 2$ factorial design on determination of differential limens.

The purpose of this note is to comment briefly on this technique and show how its use resolves the seeming paradox posed by Webb and Lemmon. The principle underlying this procedure is that, when a test of functional relationship is desired, the between SS with $k-1$ df may be analyzed into $k-1$ orthogonal comparisons, each associated with a particular aspect of the regression line. Each of these components may be tested for significance independently. The first component isolated is that due to linearity, the second is the quadratic component, the third the cubic, etc. The df associated with each of these is 1, and the individual MS's are tested against the within MS for significance.

Cases I, II, and V from Webb and Lemmon's original article will be discussed with reference to this

method. Cases III and IV have to be eliminated since it is necessary, when using orthogonal polynomials, either to assume that observations are taken at equal intervals along the X-axis, or else the actual values on the X-axis must be known. For Case I there is no problem since only two sets of observations are taken. In Case II Webb and Lemmon assume that their C group is midway between groups A and B, so equal intervals are present here. For Case V inspection of the graph seems to indicate that the four groups are spaced at equal intervals, and that is the assumption made here. For Cases III and IV it is obvious that the groups are not spaced at equal intervals, so these cannot be discussed.

To illustrate the procedure and its interpretations, the various groups were assigned arbitrary summation values (the results would have been the same no matter what values were assigned), and analysis of variance tables were constructed which yield F ratios identical with those given by Webb and Lemmon. For Case I, ΣA was assumed to be 11 and ΣB was

assumed to be 33. With 11 Ss in each group it was simple to determine the SS and MS for the between-groups variance. Since $F = 4.35$ by Webb and Lemmon's definition, it was possible to work back and determine what the within MS and SS had to be to yield this F ratio. The same procedure was followed for Cases II and V. For Case II, $\Sigma A = 11$, $\Sigma B = 33$, and $\Sigma C = 22$. For Case V, $\Sigma A = 11$, $\Sigma B = 11$, $\Sigma C = 33$, and $\Sigma D = 33$. Then the between SS for Case II was analyzed into its linear and quadratic components, and the between SS for Case V was analyzed into its linear, quadratic, and cubic components. The MS's were then computed and were tested against the within MS. All these data are illustrated in Table 1. The values which are not in parentheses are the between and within sources of variance and are identical with Webb and Lemmon's data. All the data which are enclosed in parentheses are the results of the orthogonal polynomial analysis.

When Case II is analyzed in this manner, it is seen, though the overall F test is insignificant, that when

TABLE 1

ANALYSIS OF VARIANCE OF CASES I, II, AND V FROM WEBB AND LEMMON'S
HYPOTHETICAL DATA (7)

Data not in parentheses are from Webb and Lemmon. Data in parentheses are the results of the orthogonal polynomial analysis.

Case	Source	df	SS	MS	F
I	Between	1	22	22	4.35
	Within	20	101.20	5.06	
II	Between	2	22	11	2.17
	(Linear)	(1)	(22)	(22)	(4.35)
	(Quadratic)	(1)	(0)	(0)	
	Within	30	151.80	5.06	
V	Between	3	44	14.67	2.90
	(Linear)	(1)	(0)	(0)	
	(Quadratic)	(1)	(44)	(44)	(8.70)
	(Cubic)	(1)	(0)	(0)	
	Within	40	202.40	5.06	

the linear and quadratic components are separated all the between variance is explained by the linear regression (which would have to be true since Webb and Lemmon assumed that the mean of their C group fell directly on the grand mean). Since there are 1 and 30 df to test the linear MS, this linear regression is more significant than that for Case I where the identical F ratio is found but is tested against 1 and 20 df. (Case I may also be considered to be a test of linearity, since the linear SS here is identical with the between SS.) This also follows logically since we would place more confidence in a linear curve determined by 3 points than we would in a curve determined by only 2 points.

For Case V it may be seen that a highly significant quadratic relationship exists between the two variables. None of the regression is explained by the linear or cubic elements; the quadratic component accounts for all the between variance. Though in this case the over-all F is found to be barely significant at the .05 level, a more accurate interpretation of the experiment is obtained from this further analysis.

This type of statistical analysis would appear to be profitable in psychological research. As Webb and Lemmon indicate, the situation where a functional relationship exists between the independent and dependent variables is a fairly common one. Kogan (4) in his recent review of variance designs in psychological research points out that the procedure of fitting orthogonal polynomials will frequently furnish information and answers to questions which cannot be obtained by an overall F test to treatment means. However, the fact that Kogan only cites one psychological study which has used this method—that by Johnson and Tsao (3)—would indicate that this technique has not been widely used by psychologists.

In summary, the author would like to point out that a logical analysis of the experimental situation must precede the selection of a statistical technique. In the case where E suspects that a functional relationship exists, the over-all F test of treatment means is not the best test of this relationship. The use of the method of orthogonal polynomials, however, will permit E to make an exact test of his hypothesis.

REFERENCES

1. DIAMOND, S. Comment on "A qualification in the use of analysis of variance." *Psychol. Bull.*, 1952, 49, 151–154.
2. FISHER, R. A. *Statistical methods for research workers.* (11th Ed.) New York: Hafner, 1950.
3. JOHNSON, P. O., & TSAO, F. Factorial design in the determination of differential limen values. *Psychometrika*, 1944, 9, 107–144.
4. KOGAN, L. S. Variance designs in psychological research. *Psychol. Bull.*, 1953, 50, 1–40.
5. PATTERSON, C. H. Note on "A qualification in the use of analysis of variance." *Psychol. Bull.*, 1952, 49, 148–150.
6. SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
7. WEBB, W. B., & LEMMON, V. W. A qualification in the use of analysis of variance. *Psychol. Bull.*, 1950, 47, 130–136.
8. WEBB, W. B., & LEMMON, V. W. A sequel to the notes of Patterson and Diamond. *Psychol. Bull.*, 1952, 49, 155.

TEST OF SIGNIFICANCE FOR A SERIES OF STATISTICAL TESTS

JAMES M. SAKODA, BURTON H. COHEN, AND GEOFFREY BEALL

University of Connecticut

The problem of evaluating a series of statistical tests (e.g., t's, F's, χ^2's) has recently received the attention of psychologists. The general approach to the problem is to set the significance level (p) at .05 or .01 and to find the chance probability of obtaining at least n significant results. This is done by expanding the binomial, $(p+q)^N$, where $q = 1 - p$ and N is the number of significance tests made. This procedure is applicable when prediction of the outcome of a statistical test is independent of any other tests in the series (8). Wilkinson (14) has published tables for p at the .05 and .01 levels showing the probability of obtaining n or more significant statistics out of N calculated statistics. Wilkinson's tables, however, only run to $N = 25$. Brožek and Tiede (1) in a subsequent article suggested that when Np is equal to or larger than five it is possible to employ the normal curve approximation to the binomial distribution. To do this, the critical ratio is calculated by means of the formula

$$ CR = \frac{|n - M| - .5}{\sigma} $$

where $M = Np$, $\sigma = \sqrt{Npq}$, and $-.5$ is the correction for discontinuity. The probability value is obtained from the table of area under one tail of the normal curve. To use this normal curve approximation, however, N must be 100 or larger when p is taken as .05, or 500 or larger when p is taken as .01, since it will be remembered that Np must be equal to or larger than five. Between

Wilkinson's tables and the normal curve approximation there is a gap which we feel should be filled in.

The graphs which we provide here run to $N = 100$ for $p = .05$ (Fig. 1) and $N = 500$ for $p = .01$ (Fig. 2) and can be used when the normal curve approximation is not applicable. To use the graphs, the .05 or the .01 level of confidence is selected and the number of calculated statistics (N) and the number of significant statistics (n) at the chosen level of confidence are counted. The chance probability of obtaining at least n out of N statistics can be read off the graph for values between .001 and .50. N has been plotted on a logarithmic scale, and this fact should be taken into account in interpolating for values of N. For example, for $n = 7$, $N = 60$, and $p = .05$ chance probability can be read from Fig. 1 as lying between .05 and .01. One would conclude that it is not probable that obtaining seven significant results out of 60 was due to chance alone. On the other hand, there is still the possibility that several of the seven significant statistics might have occurred by chance alone.

The current practice of tabling critical values of t's, F's, χ^2's at the .05 and .01 levels of confidence makes the counting of significant statistics at these levels and the use of the binomial distribution the most suitable approach to the problem of testing the significance of a series of statistical tests. However, there are two types of situations in which our graphs will be inadequate. The first is one in which the level of signifi-
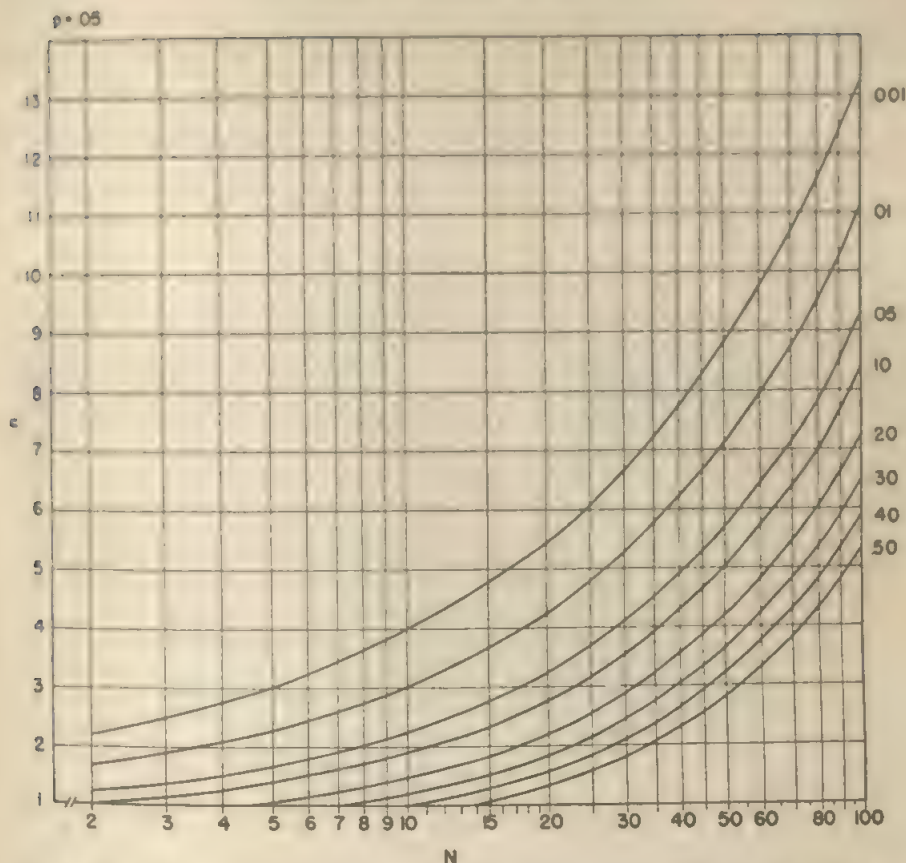
Fig. 1. Chance Probability of Obtaining at Least n Statistics Significant at the .05 Level from N Calculated Statistics

cance one desires to adopt is not .05 or .01, but, for example, .10 or .001. There are at least three ways of handling this situation. The first is to consult tables of binomial distributions. The U. S. National Bureau of Standards (16) has published tables for p ranging from .01 to .50 and N's up to 50, and Romig (12) has published tables for N's from 50 to 100. A second method is to calculate the desired binomial distribution directly, using a calculating machine and a convenient working formula (9, pp. 22–23). A third method is to use the Poisson distribution as an approxi-

mation to the binomial distribution (4). When Np is taken to be less than five (the range within which the normal curve approximation is not applicable) and p is taken to be not larger than .10, the approximation of the binomial distribution by the Poisson is fairly good even for values of N as small as two. We have found that with these restrictions the largest absolute error in calculating cumulative probabilities is .02, and in the critical area of cumulative probabilities of .10 or less the error is not larger than .012. Soper (13), Molina (10), and Hartley and Pearson (7)
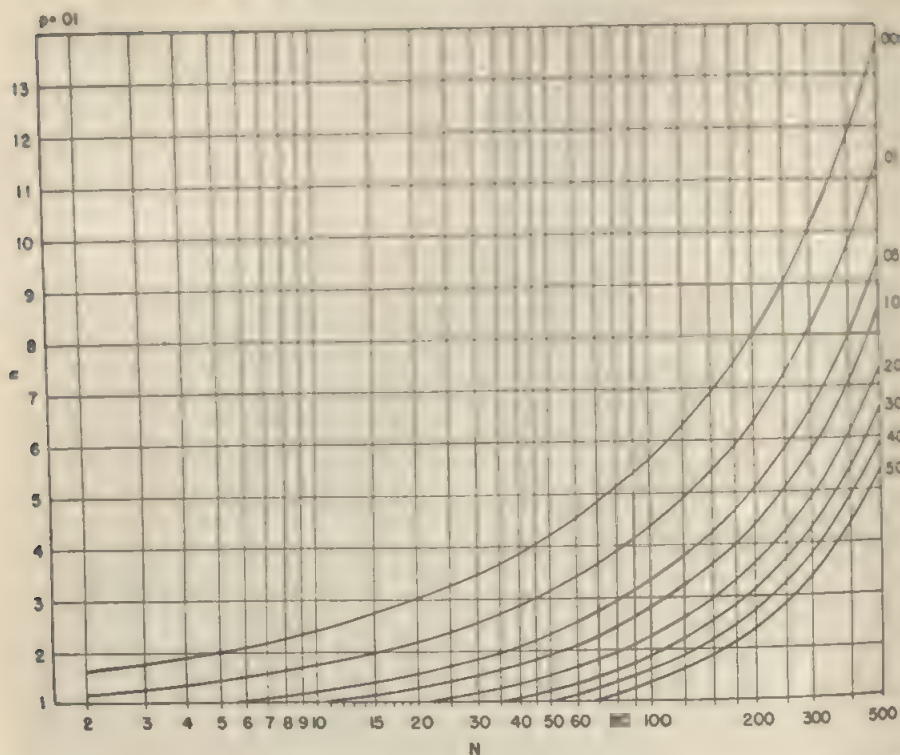
FIG. 2. CHANCE PROBABILITY OF OBTAINING AT LEAST n STATISTICS SIGNIFICANT AT THE .01 LEVEL FROM N CALCULATED STATISTICS

have published tables of Poisson distributions, and Dixon and Massey (3) include an abridged Poisson distribution in the appendix of their book.

A second type of situation in which our graphs will not be adequate is one in which exact probabilities are calculated for a number of significance tests and a sensitive test of over-all significance of the series is desired. Fisher (5) offers a method of combining exact probabilities from a series of tests of significance. His test is based on the formula for the probability of a chi square with two degrees of freedom. Using this formula it is possible to transform p values to chi-square values. Using common logarithms,

$$\chi^2 = 2 \cdot 2.302585(-\log_{10} p).$$

Independent chi squares and their degrees of freedom (two for each chi square) can be summed and these sums referred to a chi-square table for a combined probability value. For an example which is worked out, interested readers are referred to Fisher (5, pp. 99–101) or to the article by Jones and Fiske (8).

To use this method exact probabilities of t's, χ^2's, and F's are needed. Good approximations of exact probabilities can be obtained by linear interpolation in available tables of t and χ^2 by first expressing probabilities in natural or common logarithms. Since Fisher's chi-square technique calls for exact probabilities expressed in logarithms, the values found by interpolation need not be transformed to

deer antilogarithm equivalents. The same procedure applies to *F*, with the additional step of transforming *F* to square root of *F* before making the linear interpolation. The usual published tables of *F* do not allow for interpolations for *p* values below .05. However, Hald (6) in his *Statistical*

Tables and Formulas includes *F* distributions for critical *p* values of .10, .30, and .50. Exact *p* values for *F*s can also be calculated from the incomplete beta function tabled by Pearson (11), and readers are referred to Burke's (2) article explaining this procedure.

REFERENCES

1. BROŽEK, J., & TIEDE, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, **49**, 339–341.

2. BURKE, C. J. Computation of the level of significance in the *F* test. *Psychol. Bull.*, 1951, **48**, 392–397.

3. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.

4. FELLER, W. *An introduction to probability theory and its applications.* I. New York: Wiley, 1950.

5. FISHER, R. A. *Statistical methods for research workers.* New York: Hafner, 1948.

6. HALD, A. *Statistical tables and formulas.* New York: Wiley, 1952.

7. HARTLEY, H. O., & PEARSON, E. S. Tables of the χ²-integral and of the cumulative Poisson distribution. *Biometrika*, 1950, **37**, 313–325.

8. JONES, L. V., & FISKE, D. W. Methods for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.

9. KENNY, J. F. *Mathematics of statistics,* II. New York: Van Nostrand, 1941.

10. MOLINA, E. C. *Poisson's exponential binomial limit.* New York: Van Nostrand, 1949.

11. PEARSON, K. (Ed.) *Tables of the incomplete beta-function.* London: Biometric Lab., University College, 1934.

12. ROMIG, H. G. *50–100 binomial tables.* New York: Wiley, 1952.

13. SOPER, H. E. Tables of Poisson's exponential binomial limit. *Biometrika*, 1914, **10**, 25–35.

14. WILKINSON, B. A. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

15. *Tables of the binomial probability distribution.* National Bureau of Standards, Applied Mathematics Series 6, Washington, D. C.: U. S. Government Printing Office, 1950.

COMMENTS ON SEEMAN'S OPERATIONAL ANALYSIS OF THE FREUDIAN THEORY OF DAYDREAMS

RICHARD A. BEHAN AND FRANCES L. BEHAN

Michigan State College

While the present writers are in wholehearted agreement with Dr. Seeman's remark that formal operational analysis is an indispensable prerequisite to empirical investigation, it is also true that the results of the analysis, once it is completed, need to be criticized. After one has abstracted the formal structure of a theory, it is necessary to subject this formal structure to logical criticism.

The first criticism is concerned with the restatement of the Freudian assertion that daydreams are wish-fulfillments. The actual restatement is as follows: *"The emission of a daydream is functionally related to a specific type of demand (wish), the relation being such that whenever an instance of such and such a daydream is observed, it is required by the theory that an instance of a specified corresponding demand (wish) must be identified by a suitable objective operation"* (2, p. 377). Dr. Seeman then goes on to say: "It seems clear that, so stated, the theory really requires the occurrence of identifiable, lawful patterns of *demand-daydream* covariation." The point to be made here is that this restatement is not a statement in theory at all; it is, rather, metatheory. That is to say, the restatement is a statement *about* theory, not a statement *in* theory (4). If the quoted statement were actually theory, it would be an assertion of a *specific* functional relation. *There are many different functional relations which fit the description given in the quoted statement.*

Second, the theory does not require "the occurrence of identifiable, lawful patterns of *demand-daydream* covariation." The theory *is* the set of statements which assert "identifiable, lawful patterns of demand-daydream covariation." Every empirical theory consists of statements which assert identifiable, lawful patterns of construct phenomena covariation. This is required *of* the theory, not *by* the theory on purely methodological considerations.

Third, on page 377, Dr. Seeman asserts: "What is crucially important here is the understanding of the *contingent notion of frequency*, which lies buried in this analysis of the *meaning* of the concept of wish-fulfillment." Then on page 379, he asserts that Q— the hypothesis that is predicted by the theory—is a predicate. Now, a predicate, with its argument, is a two-valued constant (not a many-valued functor); it is either true or it is false. It is not possible, with the predicate Q analyzed as it is, to predict anything about the frequency of daydreams.

It is well to add in passing that Dr. Seeman is not the only psychologist who has confused statements about theory for statements in theory in this respect. There is, as it were, a great deal of precedent in modern psychological "theorizing" for calling statements of this sort theory.

The next point for consideration concerns the inductive leap from $(P \supset Q) \cdot Q$ to (P). According to footnote 7 (2, p. 378) the symbolized statement is "an extremely elementary application of symbolic logic." This is not the case. The symbolized statement is an extremely elementary

question excluded in symbolic logic —known as the fallacy of asserting the consequence (1, pp. 7–8).

The reasons that we wish to exclude formal fallacies from our theories may be summed up as follows: A fallacy is a false form of statement. It is a well-known theorem in symbolic logic that a false statement implies any statement (3, p. 104). Now, it is the case that statements asserting both sides of every question about which any theory makes an assertion are included in the class of all statements (the class including every statement). Therefore, the theory which contains even one fallacy will predict both sides of every question about which it contains an assertion. There are three results of this state of affairs which are worth mentioning: (a) The theory is never wrong—it always predicts what is in fact the case, along with what is not the case. (b) The theory makes no unequivocal assertion, and the theorist must always wait until after the fact to find out what his theory would predict. Thus, prediction is always of an *ad hoc* nature. (c) The theory is useless. Any procedure which purports to be the result of the use of the theory could proceed just as well without it.

The next point for consideration concerns the discussion of the sentence $[(P \cdot P'_{1 \ldots k}) \supset Q_{1 \ldots k}] \cdot [Q]$.

On page 379 Dr. Seeman states " . . . in those isolated instances . . . where the confirmation conditions indicate that $\sim Q$ is the case, the indication would be for a re-examination of P' before P." With this point of view we disagree. Whenever a statement derived from an empirical theory is disconfirmed the theory itself is denied. The only way out is to show that the theory does not predict the state of affairs that was disconfirmed.

With the help of symbolic logic it is easy to show that it is P that is false when one discovers a disconfirmation of the theory. There are two cases, namely: (a) P' asserts a fact; (b) P' asserts an assumption (2, p. 379). Consider the first case: If it were true that P' asserts a fact, then P' cannot be false. After all, if P' asserts an actual state of affairs, it can only be true. The theory asserts that $(P \cdot P')$ implies Q. We find $\sim Q$; therefore, by Modus Tollens, we deduce $\sim (P \cdot P')$. But P' is true; therefore, P must be false (1, p. 25).

Consider the second case: P' asserts an assumption; i.e., P' is an instance of P. The logical model for this situation is the reference formula known as Specification (Spec., see 1, p. 354). Applied to our situation Spec. asserts that P is a universal assertion that implies P'. It goes almost without saying that the universe of discourse here is the circumscribed universe in which the theory under consideration applies. Since P implies P', and we have $\sim P'$, we may, by the reference formula known as Modus Tollens, deduce $\sim P$.

It is thus seen that the occurrence of a nonconfirmation $(\sim Q)$ leads in every case to the denial of the theory. *There is no way to save a theory if it actually predicts wrong; one can only change it.*

The present writers hope that the reader will not feel that they disapprove of what Dr. Seeman has tried to do. On the contrary, it is felt that Dr. Seeman has accomplished two important things: (a) He has provided the first (to our knowledge) attempt to demonstrate the logical form of the Freudian theory of daydreams. (b) He has stated the results of his analysis in such a way that one can unequivocally determine its logical characteristics.

The present comments will be closed with a few words about the

implication of Dr. Seeman's analysis for the Freudian theory of daydreams. If Dr. Seeman's analysis is correct, then all of the remarks which were made earlier, with reference to theories which contained fallacies, are applicable to the Freudian theory of daydreams. These remarks were: (*a*) The "theory" predicts both sides of every question. (*b*) The "theory" is of an *ad hoc* nature. (*c*) The "theory" is useless; any procedure which purports to follow from the "theory" could proceed just as well without it.

REFERENCES

1. COOLEY, J. C. *A primer of formal logic.* New York: Macmillan, 1949.

2. SEEMAN, W. The Freudian theory of daydreams: an operational analysis. *Psychol. Bull.*, 1951, 5, 369–382.

3. WHITEHEAD, A. N., & RUSSELL, B. *Prin-* *cipia mathematica.* Vol. 1. Cambridge: Cambridge Univer. Press, 1925.

4. WOODGER, J. H. The technique of theory construction. *Int. Encycl. unified Sci.*, 1939, 2, No. 5.

Received March 16, 1953.

REPLY TO THE BEHANS

WILLIAM SEEMAN
Mayo Clinic

In connection with the Behan paper, I shall confine myself to a few brief comments.

1. The Behans' doctrinaire statements about theory suggest a confusion of logical fact with methodological decision, especially with the notion of what Reichenbach calls "entailed decisions" (5, pp. 11–15). The impression conveyed in their paper that their remarks on theory are in accord with Woodger (6) is also erroneous. Actually, they are in disagreement with Woodger; and the "specimen theory" he presents (and which he therefore, presumably, considers acceptable as "theory") would be ruled out by the criteria set out in the Behan paper.

2. It is not true that I have committed the fallacy of asserting the consequent. Given "If '*P*' then '*Q*' and assert '*Q*' " it would be asserting the consequent if *and only if* I were to say "assert '*P*' to be true on the basis of '*Q*.' " But my paper does not say that; what it says specifically is "inductive leap to '*P*,' " and this on the assumption that it would be understood as a convenient shorthand for something like "accept '*P*' provisionally as a proposition in the larger theory until such time (if any) that there is sufficient evidence to controvert it and render it useless in the theory." To call this "asserting the consequent" is to say either that all experimental procedure leads to this fallacy or to assert a stricture against all empirical science. And this, I presume, is what MacCorquodale and Meehl have in mind when they write "All scientific hypothesizing is in the invalid 'third figure' of the implicative syllogism" (4, p. 97). What purpose, after all, can an experiment serve if, from the results, one is forbidden to make inferences?

3. With respect to the deductions which the Behans characterize as "impossible," the most effective refutation, it seems to me, lies in the simple fact that the deductions have been done. To carry out an experimental test of this I used a group of eight professional logicians and six

experimental psychologists. They are in agreement in performing the "impossible" deductions.

4. The Behans obviously have confused material implication with logical implication.[1]

5. Their use of the formula Spec. is erroneous. This formula is not in the sentential calculus (e.g., $P \supset Q$), but in the logic of quantifiers, e.g., $x(Fx \supset Gx)$; on this, see Quine (3, pp. 17–18, ch. 2).

6. The best comments on the "proof" of the consequences of nonconfirmation for any theory are to be found in Ayer and Cohen and Nagel.[2]

7. The Behan paper is guilty of invalid inference. Specifically, in the final paragraph their conclusion would be formally valid if *and only if* there were a true additional premise "and everything we have said is factually true and formally correct."

[1] This confusion is evident in their equating material implication with prediction, after stating the "well-known theorem." It is not true that a false statement *logically* implies any statement, and hence it does not predict at all. The "well-known theorem" is a theorem in the calculus of propositions and it states that a false statement *materially* implies any statement. To equate this with prediction leads to absurdities which I shall demonstrate in a moment. The distinction between these two kinds of implication is given by Cohen and Nagel somewhat as follows: Logical implication "involves no assumption as to the factual truth or falsehood of either of two propositions, but only that they are so connected by virtue of their structure . . . that it is impossible for the implicating proposition to be true and the implied one to be false" (2, p. 127). It is quite different with

material implication, which is "the name we give to the fact that one of a pair of propositions happens to be false or else the other happens to be true" (2, p. 128). Of material implication Quine says: "This relation is so broad as not to deserve the name implication . . . " (3, p. 29).

The most effective way of demonstrating the consequences of this confusion is to exhibit its absurd results. Allowing the "well-known theorem" to formulate a prediction would lead to permitting a statement like the following: "Two plus two equals five predicts that Sacco and Vanzetti were executed for murder and Alfred Smith was defeated for the presidency in 1928 predicts that the base angles of an isosceles triangle are equal" (2, p. 127). Readers conversant with symbolic logic will have noticed that this is the same error which I committed in my 1951 paper. Had the Behans singled this out as a major error and confusion I should have had no choice but agreement. A second error was my failure to realize that the sentential calculus does not provide the resources for the formulation of a psychological theory (3, pp. 17–18, 65–71). This same error is also perpetuated in the Behan paper, which accepts the sentential calculus as the basis for argument.

[2] On this point Ayer and Cohen and Nagel are quite explicit. Ayer writes that, in the case of a nonconfirming event, "we may conclude that the [theory] is invalidated by our experiment. *But we are not obliged to adopt this conclusion* [italics mine]. If we wish to preserve our [theory] we may do so by abandoning one or more of the other relevant hypotheses" (1, p. 94). Cohen and Nagel take the same position: "The logic of the crucial experiment, therefore, is as follows: If H_1 (theory) and K (assumptions) then p_1. But p is false; therefore *either* H_1 is false *or* K (in part or completely) is false" (2, p. 220). Readers who may wonder how it was so "easy" to demonstrate the reverse of this "with the help of symbolic logic" will find the answer in the previous paragraph; i.e., the Behans applied a formula from the calculus of quantifiers to a case in the sentential calculus.

REFERENCES

1. AYER, A. J. *Language, truth and logic.* London: Gollancz, Ltd., 1948.
2. COHEN, M. R., & NAGEL, E. *An introduction to logic and the scientific method.* New York: Harcourt, Brace, 1934.
3. QUINE, W. V. *Mathematical logic.* Cambridge: Harvard Univer. Press, 1951.
4. MacCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95–107.
5. REICHENBACH, H. *Experience and prediction; An analysis of the foundations and the structure of knowledge.* Chicago: Univer. of Chicago Press, 1938.
6. WOODGER, J. H. The technique of theory construction. *Int. Encycl. unified Sci.*, 1939, 2, No. 5.

SPECIAL REVIEW

AN EVALUATION OF THE ANNUAL REVIEW OF PSYCHOLOGY (VOLUMES I-IV)[1]

LYLE H. LANIER
University of Illinois

One consequence of the postwar expansion of professional psychology—as science, as education, and as practice—has been an increasing volume of literature in all its branches. Beset by growing scientific specialization and professional diversification, psychologists individually have had decreasing time and competence to give to the assimilation of this profusion of publication. The important functions of review, evaluation, and systematization have lagged seriously behind the accumulation of articles, monographs, and books. Whether or not this material really represents a single universe of scientific discourse or can be transformed into a unified conceptual structure have been questions of mounting concern to psychologists.

Into this state of growing confusion the *Annual Review of Psychology* was introduced in 1950. It "was conceived as a supplement to, rather than a duplication of, other publications presenting abstracts or relatively long-term reviews of psychological literature . . . to present critical appraisals of current research and theory in psychology on an annual basis in the case of the most active and general fields and on a bi-

ennial basis for fields of lesser activity or scope." Four volumes have now appeared, edited by Calvin P. Stone with Donald W. Taylor as associate editor. The purpose of the present review is to evaluate this series as a whole, and to comment in particular upon Vol. 4.

GENERAL APPRAISAL OF THE SERIES

Although the *Annual Review* is widely known among psychologists, there are probably many who do not know in detail what the series has contained and who the contributors have been. Both for this reason and for its uses as a framework for the present review, a summary of information concerning all four volumes is presented in Table 1.

Size and bibliographic scope. The *Annual Review* has grown from 330 pages in 1950 (including indices) to 485 pages in 1953—an increase of about 47 per cent. This increase in length has roughly paralleled the number of references cited. In 1950 there were 1,594 titles in the eighteen chapters, while in 1953 the number was 2,003 for nineteen chapters. The gross increase was 38 per cent, while the average increase per chapter was about 25 per cent (from 89 to 111).

There are interesting differences among the authors in the number of references cited, both for different reviews of the same topic and for reviews of different topics. For example, Sears in 1950 listed only 39 articles on personality, while Bron-

[1] STONE, CALVIN P., & TAYLOR, DONALD W. (Eds.) *Annual review of psychology.* (4 vols.) Stanford, Calif.: Annual Reviews, Inc., 1950, 1951, 1952, 1953. Two of these volumes have been previously reviewed in this JOURNAL: Vol. 1 by H. H. Kendler in 1951 (pp. 159–161) and Vol. 3 by M. H. Marx in 1952 (pp. 657–660).

TABLE 1

Summary of Contents of *Annual Review of Psychology*

Topic	Volume 1			Volume 2			Volume 3			Volume 4		
	Author	Pp	Refs	Author	Pp	Refs	Author	Pp	Refs	Author	Pp	Refs
Devel...al	Jones & Bayley	8	38	Barker	22	69	Nowlis & Nowlis	28	136	Harris	30	160
Learning	Me...on	22	101	Barton	22	90	Harlow	26	121	Underwood	28	103
V......	Bartlett	18	169	Chapanis	20	96	Hobson	30	120	V......	30	154
Hearing	Newman	22	105	Weaver	14	94	Licklider	20	87	Lawender	22	92
Senses, chem.	Crozier	16	62	Pfaffmann	16	88	Wenzel	26	172	R......	26	61
Ind...lid	Thorndike	18	89	Tyler	18	91	Humphreys	20	79	Anastasi	70	139
Personality	Sears	14	39	MacKinnon	24	73	Eysenck	24	104	Bronfenbrenner	26	111
Social	Bruner	32	101	Katz	36	101	Smith	90	113	Newcomb	32	117
Industrial	Shartle	22	128	Bellows	20	98	Brown & Ghiselli	28	126	Harrell	24	140
Comparative & physiolog...al	Hebb	16	67	Deese & Morgan	24	136	Nissen & Semmes	28	157	Hess	16	71
Ab......										Neff	18	107
....	Cameron	18	142	Taub	22	216	Zubin	22	68	White	22	75
C...... ...rap.	H. Hunt	14	97	Cantlman	20	104	Margaret	18	196	R...ter	22	115
C...... ...terapy	Snyder	14	109	Pepinsky	22	117	Rosey	30	79	Sanford	26	68
Educat......	Cronbach	20	87	Stroud	24	154	Elmgren	28	104	Carter	20	129
C...... ...sgn.	Beadle	12	65	Stott	12	60				Williamson	18	99
C...... ...terapy	Bordin	10	43	Pepinsky	18	80	Gilbert	30	120			
Statist. design	Grant	20	113	Edwards	18	67	McNemar	10	33	Mosteller	28	84
Process......	Johnson	14	39									
Gerontology				Shock	18	146						
Motivation												
Spec.disabil.							Mowrer	20	55	Cobb	26	138
Theoret. psych.										Bergmann	23	40

fenbrenner cited 111 in 1953. By contrast, Cameron mentioned 142 references in his 1950 discussion of abnormalities of behavior, while White listed 75 in 1953. Such variations are, of course, very difficult to interpret since authors differ greatly in the extent to which evaluation or even individual mention is made of numbered references.

The number of references cited in Vol. 4 is roughly 25 per cent of the number of entries in the *Psychological Abstracts* for 1952. The significance of this percentage is by no means clear, however, since the *Psychological Abstracts* cites many informational and "professional" contributions which would not properly come within the purview of the *Annual Review*. It also includes many articles in related fields which would not normally be included in an annual review of psychology. Moreover, the undetermined amount of duplication among the references in the *Annual Review* would inflate this gross index of "bibliographic coverage."

It is of some interest to compare quantity of literature reviewed in the *Psychological Bulletin* with that covered in the *Annual Review*. In 1950 some 1,363 titles were cited in articles and book reviews in the *Psychological Bulletin*, while the *Annual Review* listed 1,594 references. In 1953 the comparable figure for the *Bulletin* was 1,322 while that for the *Annual Review* had increased to 2,003. These numbers are not, of course, commensurable in any strict sense, partly because of the differing objectives of the two publications and the undetermined amount of duplication among the references in each of them.

Topical organization. It is to the chapters concerned with "basic" psychological science that one turns for the *Annual Review*'s "systematic psychology"—its conceptual organization of the scientific subject matter of the discipline. Such chapters comprise approximately two-thirds of each volume. The topics fall into quite heterogeneous categories. Certain chapters are focused upon classes

of "primary" dependent variables as subject matter—general modes of behavior such as seeing, hearing, learning, and problem solving. Another set of chapters is organized in terms of "secondary" dependent variables—such as indices of abilities, personality characteristics, and patterns of social behavior. A third set of chapters is concerned with different types of behaving subjects—animals, children, the aged, the disabled, the abnormal. The material on physiological psychology is partly a miscellany, but its distinctive "systematic" concern would seem to be with the influence of a set of *independent* variables upon a wide variety of behavioral functions. In this respect it is similar to those aspects of social psychology which consider the effects of social and cultural conditions upon the behavior variables serving as systematic categories for other levels of psychological analysis.

By and large, this multidimensionality of "chapter headings" is typical of the general state of contemporary systematic psychology—a state of essentially "logic-tight" compartments insofar as formal conceptual relations go. The editors of the *Annual Review* take the field pretty much as they find it, and apparently make little direct attempt to change it towards greater logical or psychological order. This judgment is not necessarily a criticism of a publication such as the *Annual Review*. It would be difficult for the editors to do otherwise than present psychology in terms of the "chapter headings," the clusters of scientists, and the classes of institutions which underlie the confused mosaic of contemporary psychology. Scientific writers probably work more easily (and willingly!) on tasks defined in terms of familiar concepts and preferred categories. A related consideration is the fact that the literature tends to be organized in these terms (for example in the *Psychological Abstracts*) and hence is more easily surveyed than would be the case if an idiosyncratic framework for systematic evaluation were established by editorial fiat. Finally, practical considerations make it virtually impossible for the editors to achieve any kind of *ex post facto* editorial coordination or integration among the several contributions. There is insufficient time between the submission of the manuscripts by the authors and the publication of the volume to accomplish much more than routine editorial emendation of the material. The various chapters cover the literature up to within six to seven months of the publication date of a given volume. To complete the writing, editing, and printing within such a period is a remarkable achievement in any case.

Nevertheless, it is unfortunate that certain focal points of contemporary psychological discussion are largely lost within the present topical structure of the *Annual Review*. Perception and motivation are examples. Volume 3 has a chapter on motivation, but none of the four volumes has had one on perception. Most of the important literature in these areas has probably been reviewed under other headings, but there have not been the continuing, unified commentaries upon the work in these fields which their importance in contemporary psychological thinking justifies. Perhaps the heterogeneous material in either of these fields could not be transformed into anything resembling a unified body of knowledge. But it is precisely because of the confusion as to the meaning and

systematic status of such terms that continuing efforts at critical methodological appraisal of their usage in research and theory would be valuable.

More extensive, independent treatment of such topics as perception, motivation, and symbolic behavior would add, of course, to the length of a volume of the *Annual Review.* If compensating reductions had to be made they might well affect the "psychophysiological" literature. Each of the four volumes has had three chapters devoted to sensory processes, whereas for perception, motivation, and thinking only the last two have appeared once each in single volumes (see Table 1). Together with special treatment of "physiological psychology," this pattern of differential emphasis overstresses psychophysiological literature at the expense of strictly "psychological" material. The physiological bias is most pronounced in the chapters on somesthesis and the chemical senses. A check of the publication media represented in the bibliographies of two of these four chapters (Vols. 2 and 4) shows that approximately 80 per cent of the references are to be found in physiological and medical publications. Less frequent review of this literature would make space available for more adequate consideration of topics of greater significance to psychology as a whole.

The reviewers of Vols. 1 (Kendler) and 3 (Marx) in this JOURNAL thought that the several chapters devoted to clinical psychology and counseling should have been condensed and integrated. A single chapter for psychodiagnosis and another integrated chapter for psychotherapy has been the usual proposal for condensing these reviews. There certainly is overlap among the problems and techniques used, respectively, in "clinical psychology" and in "counseling." Indeed, this is a prominent subject of discussion by several of the reviewers, mainly those in the counseling area who are trying to find a distinctive definition of their field. But there are significant differences between these two fields in terms of literature cited in the *Annual Review.* For example, Gilbert's single chapter on counseling methods in Vol. 3 lists only 23 of the 196 references in Magaret's chapter on "clinical" psychodiagnostics and only 11 of the 79 references in Raimy's chapter on psychotherapy. The comparable figures for Vol. 4 show even less overlap: Williamson's "counseling" chapter includes only 17 of the 115 references cited in Rotter's review of "clinical" psychodiagnostics, and only two of Sanford's 68 references in the chapter on "clinical" therapy. To a considerable extent, of course, the degree of relative bibliographic independence may have been due to deliberate effort to avoid unnecessary duplication. Nevertheless, these figures, together with substantive differences between the parallel treatments, suggest that "clinical" and "counseling" psychologists at present have less over-all professional communality than the formal resemblances of their methodologies and underlying principles would seem to indicate. It might be added that in Vol. 5 the editors plan to have a single chapter on theory and techniques of assessment, one on psychotherapy and one on counseling methods. This arrangement is a step in the direction of "integration" and yet it will allow for special consideration of whatever distinctive problems and techniques there might be in the nonclinical aspects of counseling.

Evaluative character. The pattern of

topical organization provides only the structural framework for the appraisal of the literature. Other important determiners of the success of an annual review are the selection of contributors, the formulation of the reviewing task, and constructive acceptance by the reviewers of the task defined. Concerning the contributors, the list of authors in Table 1 is evidence that the editors have in general succeeded in securing psychologists from whom outstanding evaluations of the literature would be expected. The pessimists who predicted that the *Annual Review* would fail to recruit a continuing force of first-rate contributors were wrong—so far as these four volumes go.

With respect to definition of the task, the preface to Vol. 1 states: "They [the authors] were at liberty to indicate their points of departure by citing summaries of basic work prior to the review period but were advised not to strive for mere comprehensiveness, especially in the form of a loosely integrated series of abstracts of all the current literature. On the contrary, the editorial board asked that they adopt an interpretative and evaluative approach to the literature they selected for review." Thus the editors clearly foresaw the danger that the reviews might degenerate into unorganized bibliographic digests. They appeared to recognize also that certain of the reviews in Vol. 1 had failed to avoid that danger, since they remark further: "While it appears that all fields are not equally amenable to this approach, it is hoped that future volumes of the *Review* will reflect gains in the techniques of exposition and interpretation of the desired kind."

To what extent have the authors been able to achieve the general objectives set for them by the editorial board? Have successive volumes shown improvement in the "techniques of exposition and interpretation?" Are there persistent differences among fields in the "evaluative character" of the reviews? In the attempt to answer these questions, I devised a procedure for appraising individual chapters in terms of three sets of presumably desirable characteristics: (*a*) an introductory orientation which defines the field and establishes a general methodological setting for the review; (*b*) explicit indication of the author's "systematic" approach to the field, including the rationale for his topical organization and bases of selection among references; (*c*) interpretative and evaluative comments upon the literature, including a summary assessment of over-all achievements and trends (e.g., in methodology, in empirical knowledge, in theory, in technology—as appropriate). In the effort to simplify the difficult judgmental task, the appraisals were restricted to "factual" estimates as to whether or not (for the first two criteria) and to what extent (for the third) the chapters exhibited these characteristics. As far as possible, differences in quality of writing, in technical sophistication, and in validity of argument were disregarded.

The results of these appraisals are represented by the figures in Table 2. In general, it appears that about one-third of the 63 chapters common to the four volumes lack the kind of "orientational" introduction described above. More than half of them omit any explicit discussion of the reviewer's "systematic" analysis or organization of the field. One-fourth have little or no interpretation or evaluation of the literature reviewed. These generalizations are subject, of course, to whatever quali-

TABLE 2

Results of Ratings of Chapters for "Orientation," "Systematization," and "Interpretation-Evaluation"[a]

Classification	Volume			
	1	*2*	*3*	*4*
No. of chapters without "orientation"	8	6	5	3
No. of chapters without "systematic" rationale	10	12	9	9
Extent of interpretation and evaluation[†]				
Median rating	4	3.1	5.5	4.8
No. of chapters rated 1 or 2	5	6	2	3

[a] The chapters common to all four volumes were in-
.... as .... condensed or expanded in the
last two volumes (see Table 1).
[†] The chapters were rated on a 7-point scale in which
"1" represented the minimal and "7" the maximal de-
gree of "interpretation-evaluation."

fications might be required by the unreliability of the method and by possible idiosyncrasies in this single reader's standards of judgment.

The editors' hope for improvement in future volumes—expressed in Vol. 1—appears to have been fulfilled, if the figures in Table 2 can be accepted as evidence. Clearly, Vols. 3 and 4 excel the first two in estimated degree of interpretation and evaluation, with Vol. 3 standing highest and Vol. 2 lowest in this respect. Two mitigating circumstances in behalf of Vols. 1 and 2 should perhaps be noted: (a) the writers of the later volumes could profit by seeing the earlier ones and from critical reactions to them on the part of the editors and other critics; (b) the last two volumes are considerably longer, thus presumably making more space available for evaluative commentary. The weight of these considerations is reduced, however, by the fact that individual reviews in Vols. 1 and 2 are rated just as high as those in Vols. 3 and 4. The incli-

nation and critical capability of the individual author are probably the principal determiners of the variance among reviews in evaluative character.

There might, nevertheless, be differences among fields in "amenability" to the kind of reviewing in question, as the editors have suggested. Although not conclusive evidence for such a question, my individual ratings show that the reviews of the applied fields are less "systematic" and "evaluative" than those in the strictly scientific areas. Again, however, the intragroup differences are much greater than the average differences between the groups. Certain of the reviews in clinical psychology stand as high by our criteria as those in any other field, in contrast to the reviews of industrial psychology. With such a broad range of application of techniques from so many branches of psychology, the literature in the latter field is quite heterogeneous, and many of the studies are highly "atomistic" in nature. Integrative evaluation and systematization are obviously easier to accomplish in an applied field such as clinical psychology where the material is organized more definitely in relation to general principles and theory.

Volume 4—1953

With its nineteen chapters and 485 pages (including indices), this largest member of the *Annual Review* family presents one innovation which will be welcomed by most psychologists: the inclusion in several of the bibliographies of the full titles of all references. This option was apparently available to all authors but was exercised by only seven of them. (It is disappointing to find that the chapter on special disabilities follows the

Annual Review's original practice of numbering references in the order of citation in the text.)

There are two "special" reviews: a chapter by Bergmann on theoretical psychology and one by Cobb on special disabilities. Otherwise, the organization of topics remains essentially what it was in earlier volumes (see Table 1). There are separate chapters on comparative and on physiological psychology in place of the former combined chapter, and counseling continues to have a single chapter as in Vol. 3. In general, Vol. 4 is a very impressive addition to the evaluative literature of psychology.

Child psychology. In contrast to Barker's soul-searching analysis of the uncertain and declining state of this field in Vol. 2, Harris presents a picture of buoyant empiricism, documented by the longest bibliography in the book. With so much to write about, he doesn't worry over the difficulty of defining the field rigorously—in contrast to previous reviewers. He is mainly concerned to report upon the 160 studies in the bibliography, in the form of brief summaries of method and results. There is little evaluation or systematic interpretation, and the organization of the material is poor.

Learning. Underwood's review continues the generally high level of previous reviews in this field—particularly with respect to careful interpretation and evaluation of significant studies. He excludes from consideration the host of investigations in which learning is merely involved or demonstrated, restricting the review to "work in areas with established methodology and some theoretical orientation." This means a fairly heavy concentration upon animal learning and human conditioning, an

unfortunate necessity arising from the state of the field and which Underwood probably deplores as much as anyone. Research on thinking, he notes, "continues to move at an appallingly slow pace," although "there are rumblings which portend better things to come."

Little attention is given to questions of over-all systematization of the field of learning, either by way of introductory orientation or in the main body of the review. Perhaps conceptual cross references and integration are not possible for most of these presently somewhat discrete problems, but continuing attention should perhaps be given to these desiderata on the part of reviewers in the *Annual Review* and in similar publications.

Sensory processes. Vernon's review of visual literature is the only one of the three chapters on the "senses" which presents an organizing framework for the literature to be reviewed. Licklider for hearing and Ruch for somesthetic and chemical senses get under way merely with brief indications of what their respective reviews will emphasize. Once past the introduction, however, Vernon contents herself pretty much with descriptive reporting of a wide range of literature (extending from simple threshold phenomena to the "New Look" studies of motivational factors in perception). Licklider, on the other hand, writes a highly lucid interpretative evaluation of the auditory literature. It appears to an outsider to be a balanced account of the entire spectrum of research in this field, insofar as a single year's output yields such a product. Ruch's chapter is restricted almost entirely to research on physiological mechanisms underlying somesthesis, with only about three pages on the "chemi-

cal senses." It is a well-written review, somewhat better in respect of interpretation than of evaluation.

Comparative and physiological psychology. Hess remarks about comparative psychology that it is "probably high time we stop deploring the lack of an orienting theoretical approach." He finds one in the work of European investigators, particularly Tinbergen and Lorenz. With this general orientation he proceeds to give a good account of a relatively small number of studies. Other comparative psychologists, notably Schneirla, will probably disagree with Hess's view that this recent European work is the needed "shot in the arm" for their somewhat neglected field.

Neff doesn't discuss the systematic status of physiological psychology, although he commences his review with two brief paragraphs which mention the problems of major current interest to the specialist in this field. He then proceeds to a descriptive digest of studies classified under these headings: sensory discrimination, basic drives, emotion, learning-memory-problem solving, biochemical and neuroanatomical changes in mental disease. There is almost no interpretation or evaluation.

Differential psychology and personality. Four chapters fall under these general headings: individual differences (Anastasi), personality (Bronfenbrenner), abnormalities of behavior (White), special disabilities (Cobb). All four fields are concerned with psychological differences among individuals and classes of individuals.

Anastasi stays within the psychometric framework in writing a well-organized review of the work on individual differences. The discussion is rather more on the interpretative than the evaluative side, except in

the parts concerned with the biological and social determinants of these differences. There she is critical of "hereditarian" inferences from inadequately designed research.

Bronfenbrenner's chapter is well written and it ranks high in the dimension of "interpretation-evaluation." His introductory section states clearly his "biases" and his justification for them. Recognizing that "the attempt to deal . . . with multiple, interacting relationships rather than simple static entities often leaves the theoretician with a system of ambiguous abstractions," he prefers the opportunities of the former to the narrow constraints of the latter approach.

White's chapter on abnormalities of behavior has a far more pronounced psychodynamic orientation than its predecessor by Zubin in Vol. 3. The latter stressed organic and genetic factors in his selection of references and in his general evaluative framework. White includes such material—in particular a good discussion of biochemical investigations of schizophrenia—but he emphasizes the influence of learning and life history upon abnormal behavior. He concludes with the hope that "if we can arrive at more basic biological reaction patterns and also at more basic psychological reaction patterns these two classes of events will prove at last to be truly convergent." In terms of its technical characteristics as a review, this chapter is one of the best in the book.

Cobb defines a "special disability" as any defect or disability that may occur in an otherwise normally functioning person. Her bibliography is comparatively long, but its utility for readers is greatly reduced because it is not in alphabetical order. The material reviewed is a mixture of

basic and applied research on the etiology and psychological characteristics of visual, auditory, and speech defects, together with publications on treatment and rehabilitation. The chapter is well written and is in general an excellent survey of a field not widely known to most psychologists.

Social psychology. Newcomb opens his review with an interesting innovation. By way of appraising efforts to define social psychology, he makes a comparative analysis of the content of six recent textbooks. His conclusion is that "social interaction" is the distinctive subject matter of social psychology, "that the term stands for something which can be studied at its own level." The phenomena of social interaction cannot be explained, he thinks, by "mere extrapolation of general psychological principles."

Newcomb organizes his material under two main headings: attitudes and the processes of interaction and communication—without discussing the systematic interrelationships among the problems and concepts of the two areas. The presentations of research findings are exceptionally good, although there is less critical evaluation than in Smith's chapter in Vol. 3.

Clinical psychology and counseling. Rotter sees the situation in psychodiagnosis much as Magaret did in the preceding volume: "experimental investigation of test validity emphasizes more and more the absence or inefficiency of prediction." His review is well planned, and probably has as much structural unity as a good theoretical orientation can give to an appraisal of work on a multitude of incommensurable tests. There are three broad headings: (a) general contributions to the methodology and

theory of personality measurement (b) clinical instruments; (c) research instruments. There is a good summary at the end called "Analysis of Trends." Reviewers in other fields would do well to emulate this feature.

Sanford's account of psychotherapy is not as systematic or comprehensive as Raimy's chapter in Vol. 3, although the chapter as a whole is a balanced interpretative appraisal in which clinical and scientific values are happily blended. A good report is given of the work in group therapy at the Tavistock Clinic in London, based upon personal observations over a period of several months. The longest section in the review is devoted to research on psychotherapy, which Sanford regards as a fruitful means to the general investigation of personality.

Williamson classifies the literature on counseling into (a) publications oriented towards psychotherapeutic theory and techniques, and (b) those focused upon "the choosing of occupational goals based upon the diagnosis of aptitudes and interests." He hasn't much to say about therapeutic studies, devoting most of the review to investigations of various diagnostic instruments. There is very little interpretation or evaluation.

Educational and industrial psychology. Superficially dissimilar, these two fields have many points of methodological and technological resemblance. Both involve a broad range of psychological principles and practices designed to improve the effectiveness of fundamental social functions. Both draw upon all levels of psychological science—general, individual, and social. Perhaps for this reason, neither represents a unified scientific-technological speciality—as the two chapters in Vol. 4 will attest.

For Carter, educational psychol-

... deals with school learning and its correlates. His review begins with a brief outline of the problems and conditions which further define the field. Such topics as readiness for learning, acquisition of desirable attitudes, meaningful learning, sociological correlates of learning, emotional factors in learning, and measurement problems predominate. Very little work on the learning of specific school subjects per se is reported. And there is almost no relationship between this literature and that reviewed by Underwood for the general psychology of learning. On the whole, this review is a good one, in marked contrast to that of Elmgren in Vol. 3.

The lack of evaluation in the reviews of industrial psychology has already been noted. Harrell does, however, present a fairly extensive introduction in which over-all trends are effectively discussed. Major emphasis in the body of the review goes to studies of human relations in industry. As a rule these field studies appear to be inconclusive because of such factors as inadequate samples, uncontrolled conditions, and uncertain criteria.

Statistical theory and research design. Mosteller's chapter is considerably longer than its predecessors in earlier volumes, partly because it covers a wider range of topics. Special features of the review are discussions of nonparametric statistics and ranking methods, which are being used increasingly by psychologists whose data often fail to satisfy the assumptions of the more conventional parametric methods of analysis. There is an informative section in which studies dealing with the effects upon the common parametric statistics of departure from assumptions are reviewed. The general conclusion is "that departures from assumptions

may have little effect at some points of the distribution function, say near the mean or median, but far out in the tails of the distribution errors get to be large and the model may just no longer be appropriate." He suggests that it would be useful to have empirical studies of the effect of violating the assumptions to varying degrees on this model. His attitude—not unusual in mathematical statisticians of empirical bent—seems to be that while illicit inferences are to be avoided, unnecessary scrapping of information is likewise to be minimized. Questions of the utility and limitations of various other mathematical "models" are discussed in a special section, including efforts to develop probability models for learning data, uses of information theory, and difficulties in finding a suitable model for accident proneness.

Theoretical psychology. This is Bergmann's term for the logic of psychology. As a branch of the philosophy of science, theoretical psychology in this sense is concerned with the nature and structure of psychological concepts, with the laws in which they occur, and the theories into which these laws combine. Following this definition of the field, the author outlines the essential tenets of a "philosophy of psychology" on which he thinks virtual agreement has been reached, and devotes the remainder of the chapter to a critical analysis of deviations from these precepts.[2] First, certain dissents from this position in principle are noted: (*a*) the phenomenology of Snygg and Combs as regards the derivation of concepts; (*b*) concerning laws, London's antideterminism; (*c*) Skinner's attack upon theory. After brief refu-

[2] To designate this general methodological position, Bergmann uses the terms "neobehaviorism" or "logical behaviorism."

tations of these attacks, Bergmann gets down to the main business of the chapter which is the "prophysiological" movement in theoretical psychology. This development has two foci: (a) the Meehl-MacCorquodale-Feigl argument for the use of hypothetical constructs as distinct from intervening variables, and (b) neogestaltism as represented in recent articles by Krech. Although Bergmann's analysis may not convince his antagonists, I think it is an outstanding contribution towards the clarification of the issues. It is perhaps unfortunate that he had to use so vulnerable a target as Krech for his attack upon gestaltism. But the gestaltists are mainly to blame; very few of them have ventured into the realm of formal methodological argument.

BOOK REVIEWS

SHAFFER, G. WILSON, & LAZARUS, RICHARD S. *Fundamental concepts in clinical psychology*. New York: McGraw-Hill, 1952. Pp. xi + 540. $6.00.

This book could well have been two separate publications. The authors have divided their efforts so that one half the chapters are written by one and the remainder by the other. There is not much similarity in approach or in style between the two sections. The first eight chapters are a sort of annotated bibliography covering the clinical field. The coverage is quite broad but not especially deep. The last seven chapters are largely devoted to psychotherapy and related topics.

The book should be useful in a number of ways, one of which will be as an aid to students preparing for comprehensive examinations. It includes a great deal of information and it directs students to other excellent sources.

It does not, however, live up to the promise of its title. Probably a representative group of clinical psychologists could not agree completely about which concepts in the field are really fundamental. But it is likely that most of them would agree that concepts such as psychological determinism, frustration-aggression, func-

tional autonomy, dynamic mechanisms, and partial reinforcement, to name a few, are central to the field. None of these is given more than cursory attention. Although psychoanalysis is discussed briefly in two different sections, the ego mechanisms are not included. The kindest thing to be said for the discussion of the Oedipus complex is that it is inadequate.

The book is written in the first person plural. The editorializing, which appears frequently, will be annoying to some readers and stimulating to others. The authors often criticize other clinicians for careless validation, slipshod methodology, and cloudy thinking. On the other hand, many readers are likely to take exception to their undocumented statements about the therapeutic process. In discussing therapy the authors stress the importance of such things as self-knowledge on the part of the therapist, the mastery of a variety of techniques, and special forms of therapeutic training. Such statements as "insight or uncovering therapy is, of course, the preferred treatment for most adjustive difficulties," or "reassurance is necessary in all therapeutic situations," or "the patient must uncover repressed material" will strike some

readers as dogmatic and unsupported by firm data.

It is likely that there would be no general agreement among clinicians that a book dealing with fundamental concepts in the field should include five chapters on psychotherapy. There would probably be fewer votes for the necessity of a chapter on physical and chemical therapies. The latter chapter is quite detailed down to citing dosage of Mitragol and methods of applying electrodes. Much material in the earlier chapters is abbreviated or omitted on the plea of space limitation. Judgments will differ on questions of what should have been included and what left out. Many will feel that the book overemphasizes therapy at the expense of other concepts.

One final objection will occur to some readers. The authors constantly refer to persons with behavior disorders or adjustive difficulties as "patients" who "suffer" from "illnesses." This seems to imply a fundamental concept about the nature of the clinician's problem which should be further examined. Thinking about problems is frequently affected by the words used in their definition, and many psychologists feel strongly that by calling people with emotional difficulties "sick" there are created unnecessary semantic barriers to effective action.

In a work that tries to cover as much ground as this one it is easy to find faults and to disagree with the authors' choices of material for inclusion. The authors have made a real attempt to cover the field. Readers' estimates of the success of the effort will vary with their agreement with the writers' choices of topics emphasized.

GEORGE W. ALBEE.
University of Helsinki.

KLEIN, MELANIE, HEIMANN, PAULA, ISAACS, SUSAN, & RIVIERE, JOAN. (Eds.) *Developments in psychoanalysis.* London: Hogarth Press and Institute of Psycho-analysis, 1952. Pp. viii+368. 30s.

The psychoanalytic dictum that early infancy is the crucial developmental period has been a source of as much frustration as stimulation among psychologists. While no psychologist these days seriously questions the importance of this early period, neither is any psychologist particularly satisfied with the evidence for its importance or the methods for its study. The retrospective reconstruction of infancy through adult analyses, and the description of babies' feelings and attitudes by way of adult empathy, have always left psychologists dissatisfied and impatient. On the other hand, there are no conventional, generally accepted psychological methods for interpreting the emotional significance of infant behavior. Consequently, there are many theories, little data, and deep frustrations.

The present collection of papers, in adding much to theory and little to data, may strain the frustration tolerance of the psychological reader. For the most part, the book aims to clarify Melanie Klein's theoretical position concerning early infancy. As usual, her contributions are imaginative, provocative, and controversial. A fair evaluation of the writings, however, requires criticism on two different levels.

One may begin by considering the theoretical contributions alone. Here there is little question that real advance has been made in the sharpening and extension of Klein's concepts beyond the position she maintained in *The Psychoanalysis of Children* (1932). The emphasis on Freud's

concept of death instinct, and its use as the source of basic infantile anxiety, provides a starting point for the first of two phases of infant development: the *paranoid-schizoid position*. A major defense during this early period is splitting—basically the splitting between the "good" and "bad" breast, but reflecting as well a splitting or nonintegration of the ego. It is only as the ego becomes able to sustain anxiety that the infant can adopt the more mature *depressive position*, which itself fosters defenses leading to greater ego integration. The relationship of this developmental sequence (greatly oversimplified by the reviewer) to the growth of introjection and projection represents a major contribution to psychoanalytic theory.

Not everybody will agree with this expansion of theory. Many of the points made by Klein have been for years a source of controversy between one English analytic group and certain European analysts. Too often, this controversy has ended with mild name-calling: who is the "genuine" psychoanalyst and who the impostor? Frequently enough, it has involved real questions as to the appropriateness of an extension of certain Freudian concepts to the earliest months of life.

One cannot really evaluate theoretical contributions alone, however. Internal consistency is never enough. The more closely and carefully the detailed accounts of postulated infant development are written, the more they cry for empirical test. The test is not to be found in these papers, although a suggestion as to Klein's methods of arriving at hypotheses is. The instances which confirm Klein's hypotheses, as reported in one chapter, are taken from published British studies of infant development, from

her own clinical observations, and from several sources and accounts of what babies do. As always, Melanie Klein in her observations is the sensitive, intuitive investigator from whom few meanings of infant behavior are hidden. Riviere and Isaacs add notably to her accounts. But there is no consideration of alternative hypotheses, no account of the negative case, no hint as to the path from observed behavior to intuitive interpretation to theory construction.

Even in the absence of these requisites of scientific method, the psychologist might be satisfied if the hypotheses advanced were testable. A careful survey of Klein's notions suggests that some probably are subject to test, although these are the peripheral rather than the crucial hypotheses. If one cannot study the content of the unconscious fantasy of a three-month-old infant, for example, one still may be able to vary the frustration to which he is subjected and observe the typical defenses. If one cannot be sure of the infant's attitude toward the breast as part-object, one may still examine the relative need-satisfying properties of food and attention. Although one may not yet be able to relate weaning to persecutory anxiety in the infant's thoughts, one may nevertheless look to the spontaneous play of weaned infants as an index of the significance of lost objects to the child.

It is an encouraging sign that numerous isolated articles in the literature of child development touch upon just such points as these. Today their number is increasing. Indeed, much of Klein's theorizing sounds less bizarre now than it did in 1932, largely because we are gradually accumulating from infants observational material which has been ordered in

probl... analytic terms. The lasting importance of these papers may well be not so much in the theoretical controversies they foster as in the impetus they give to the systematic observation of infants.

ANN MAGARET GARNER.
University of Illinois,
College of Medicine.

GREENE, EDWARD B. *Measurements of human behavior.* (Rev. Ed.) New York: Odyssey Press, 1952. Pp. xxiv+790. $4.75.

Readers familiar with the first edition of this book will be prepared for the comprehensive scope of the revision and its vast compilation of detailed and up-to-date information about measuring techniques of all varieties. They will be disappointed, however, in their expectations of a more tightly integrated book with a clearer and more precise exposition of basic principles and concepts of psychological measurement.

The fault may lie in the nature of the task the author set for himself. What was difficult in 1941 has become virtually impossible a decade later. Measurement and evaluation devices have proliferated rapidly in all areas of psychology and the development of their systematic rationale has not kept pace with empirical applications. Any book which attempts to encompass practically all the techniques (and little of the underlying theory) which have been or are employed in quantifying and assessing human behavior must suffer from some superficiality and lack of coherence. And the range covered in this one is formidable: from psychophysical methods to the Szondi test, from Thurstone's attitude scaling to OSS evaluation procedures and Kinsey's interviewing methods. Unfortunately, the book is also exasperatingly crammed with minutiae, often irrelevant. Even the subjects who posed for the photographs of test administrations are graciously identified by name.

The introductory chapters provide only a weak foundation for supporting this mass of material. For example, the problem of the nature of measurement is dispatched in a paragraph; the procedures for developing measuring instruments are given in the form of brief exhortations such as the following: "decide specifically what is to be measured and how; . . . secure a large number of sample items . . ."; or, "analyze the responses to each item to determine such attributes as content. . . ." The other introductory chapters deal with test nomenclature (in which a curious distinction is made between achievement, aptitude, and psychological [!] tests) and the characteristics of a "good instrument." Apart from a cursory discussion of reliability, norms, and the questionable attribute of test "uniqueness," the author never develops the general methodological considerations essential to an understanding of test construction and accurate interpretation of test results. Such questions as the extent to which psychological tests satisfy the criteria for measurement, the logic underlying the concepts and empirical determination of test reliability and validity, the selection of the standardization group, etc. are disregarded or only incidentally treated. Nor are the concepts and procedures included in the three chapters on "elementary statistics" (one of which deals with factorial analysis) effectively applied to the problems involved in evaluating and using tests. In this connection, it is regrettable that the chapters in the earlier edition on "persistent problems" were deleted.

The remaining 500-odd pages are devoted primarily to a description of specific tests and related tools. There are seven chapters concerned with tests for early childhood, individual tests of ability, group tests of ability, motor and mechanical tests, tests of special aptitude, educational achievement tests, and those developed for military personnel in World War II. In each of these chapters, representative tests are described and illustrated, scoring techniques explained, and some of the research findings on the use of the test surveyed. Many new and valuable tests developed since the appearance of the first edition have been included. There are, however, a few surprising omissions like the WISC and the Merrill-Palmer whereas some comparatively ancient tests of limited value have been retained.

The chapters on attitude-, interest-, and personality-measurement have been greatly expanded. They are preceded by a distressing introduction to dynamic theory and structure of personality. The need for such a chapter is clear since many projective tests are discussed. But the oversimplified and, in places, muddled account points up the unrealistic design of the book. Clinical tests like the Bender-Gestalt, TAT, Rorschach, and Draw-a-Person are described, for example, quite fully. Such material, however, cannot serve in lieu of training manuals, nor is the methodology sufficiently explored to give the reader insight into the complex problems related to the validation as well as the clinical and research use of these tools.

Some errors of fact and naïveté are understandable in a survey text of this scope. Still, inaccuracies like the following are too frequent: "the word trait . . . is used to refer to any physical aspect of a person . . . or to any mental aspect such as speed of reading . . . "; "the sum of the verbal test scores [on the W-B] yields a verbal MA . . . "; "this [the standard deviation] provides a method of scaling scores . . . comparable with the best physical scales . . . "; "the K score [on the MMPI] is the number of answers omitted because the client cannot say or will not choose."

The breadth of content combined with the unsystematic approach limits the audience for the book. It can probably best serve as a reference text for advanced students who wish to become acquainted with a wider variety of tests than the standard laboratory and practicum courses are able to cover. The profusion of visual illustrations, the research bibliographies, and the fairly complete classified lists of available tests and inventories should be particularly helpful for this purpose.

EVELYN RASKIN.
Brooklyn College.

BAUMGARTEN, FRANZISKA. (Ed.) *La psychotechnique dans le monde moderne.* (Psychotechnics in the modern world.) Paris: Presses Universitaires de France, 1952. Pp. xi+630. 2,000 fr.

Appropriately, these *Proceedings* of the Ninth International Congress of Applied Psychology were published in the series *Bibliothèque Scientifique Internationale* (Sciences Humaines, Section Psychologie, with H. Piéron as the chief editor), designed principally as means for acquainting the French-speaking psychologists with important developments in the field of scientific psychology. In this series were published the translations of Rorschach's *Psychodiagnostics* (together with the Böchner-Halpern volume on clinical applications of the

Rorschach test); G. H. Thomson's treatise on factorial analysis of human ability; and H. J. Eysenck's work on the dimensions of personality; and a large number of books by American authors (W. H. Sheldon's *Varieties of Human Physique*, two volumes by A. Gesell and F. Ilg, C. Wolff's *Human Hand*, as well as R. S. Woodworth's *Experimental Psychology*, C. T. Morgan's *Physiological Psychology*, and D. Krech and R. S. Crutchfield's *Social Psychology*). Among the few volumes by French authors are J. M. Faverage's book on statistical methods and a book on early child development by O. Brunet and I. Lezine.

There has been an interval of a long, eventful 15 years between the Eighth Congress of the International Psychotechnical Association (Prague, 1934) and the Ninth Congress (Bern, 1949). The editor stressed in her Introduction that the Congress has offered to the participants an opportunity to become acquainted with the present, dramatically altered features of psychotechnology.

To provide a historical perspective, Dr. Baumgarten, as the secretary general of the Association, requested reports on developments in the field of applied psychology in different countries during the war years. The communications were published under the title *Progress of Psychotechnics* (1939–1945) by A. Francke in Bern. The continued interference of political ideologies with the handiwork of the applied psychologist is reflected in the fact that the Polish delegation, in a letter of May 5, 1950, withdrew the manuscripts of papers prepared for the Congress, which they were unable to attend.

It appears as a wise policy that the *Proceedings* have not been limited to abstracts of the papers nor do they

reflect with photographic accuracy what has transpired. The papers have been frequently shortened and regrouped. Among the outstanding recent trends in psychotechnology, Dr. Baumgarten noted the concern with social problems (including human relations in industry) and the search for tests of personality.

JOSEF BROŽEK.

University of Minnesota.

LAWSHE, C. H. (Ed.) *Psychology of industrial relations.* New York: McGraw-Hill, 1953. Pp. vii + 350. $5.50.

The stated purpose of this book is "to present in a reasonably concise and non-technical form some of the content of industrial psychology which . . . might be useful to people who work with or manage other people." The reader will find that the authors have been very successful in fulfilling this purpose. New material is presented in a style readily understandable by the supervisory force and by other management personnel. An exhaustive list of references at the end of each chapter makes it possible for the reader to investigate any topic in greater detail.

The rigorous scientific approach to problems typical of the Purdue group is evident throughout the book, and the application of the experimental method to industrial relations is documented with practical examples.

One example of the critical attitude exercised by the authors in evaluating the studies cited can be seen in the chapter on employee supervision, in which the results of a nationwide survey of foremen are reported. The original article indicated that many of the foremen felt they were not posted on company

policies and that union shop stewards and others had usurped some of their rights. The authors of this book point out that their conclusions cannot be justified by the basic data and that only a minority of foremen held these unfavorable views.

Three chapters of the book are devoted to employee and group relations. These chapters are outstanding and should prove most helpful to management personnel. The discussion of the informal group in the work force will enable foremen to understand the reasons for its existence. When the supervisors realize that the informal group insures the individual personal recognition in a highly impersonal system and supplements the formal organization structure they will be better able to carry out their duties.

Naturally, no book contains all the material the reader thinks should be included since there are practical limits of size which must be observed. Throughout this book a few chapters would have been strengthened by the addition of new material or a more thorough discussion of the material presented. The chapter on motivation and discontent in industry is somewhat oversimplified even for the nonpsychological reader.

It is implied that the industrial psychologist is able to determine the motivation for an employee's behavior by examining the work and home environment. If the reader is not aware of the effort required and the limitations of the present state of the art, he may well expect overnight changes in the work force and subsequently be disappointed when they do not occur.

In the chapter on employee counseling it is suggested that in some organizations the first-line supervisor will have to double as employee counselor. This is undoubtedly true,

but not enough attention is devoted to the fact that it would be a difficult task for the supervisor to function in this dual capacity. The supposition that the supervisor can set aside his interests and activities for the time he acts as the employee counselor is questionable.

Unfortunately, a major failing of this book is the fact that the authors ignored union-management relations under the Taft-Hartley Act. The provision for an enforced "cooling off" period under this Act has very definite implications for industrial relations. The worker, the union shop steward, and the foreman must try to "carry on" as usual during a period when negotiations would have reached a climax under other conditions. Worker-foreman relationships should also have been considered in the prestrike and poststrike periods.

This book represents an effort to fill a gap where no adequate literature previously existed. It will be very helpful to supervisory personnel and will enable many people who do not have specialized training in this area to understand and use industrial psychological techniques. It is further suggested that this book would prove valuable as a second text in management training courses.

JOSEPH W. WISSEL.
Dunlap and Associates, Inc.

POWDERMAKER, FLORENCE B., & FRANK, JEROME D. *Group psychotherapy: studies in methodology of research and therapy.* Cambridge: Harvard Univer. Press, 1953. Pp. xv+615. $6.50.

The meat of this book lies in the numerous detailed descriptions of critical events arising in an extensive series of group therapy sessions with neurotic and schizophrenic veterans, events which have been maturely reflected upon and used as a basis for

helpful generalizations about the process and technique of group therapy. The authors and their collaborators are sensitive observers, conscientious reporters, and sensible commentators. The trimmings of the book lie in the research data reported, which are nice to have but not nearly so nourishing as the analyses of group meetings and the excellent commentaries thereon.

The report is based on a group effort lasting over a period of two years. Some 120 neurotic patients of a mental hygiene clinic and some 170 chronic schizophrenics in a psychiatric hospital were treated by 19 different psychiatrists working with a total of 27 groups. A research team of eleven persons, including psychiatrists, psychologists, and social workers, observed and recorded the meetings and then worked together to make communicable sense of the results of their experiences. An average of five patients attended the clinic groups. The groups of psychotics were more than twice as large, a circumstance reported as favoring therapy with these severely disturbed people. One group, and a fascinating one, was the entire population of a ward, about 85 patients. Most of the therapists were inexperienced in working with groups at the start of the project, and though they seem to share confidence in catharsis and interpretation as sources of gain in therapy, their newness to the group situation ensured considerable variability in approach. Groups varied in size and in rationale of composition; some groups were open, others were closed; duration of therapy varied from 8 to 175 meetings. In the neurotic group most patients received individual therapy concurrently. Such sources of variability are splendid for generating hypotheses but impose tangible limitations on their verification. An observer sat with each group and was responsible in collaboration with the therapist for recording what went on. Tape recorders were used as an aid to accuracy but not as a source of verbatim transcripts. The accounts of sessions and of critical events were made more graphic and complete by descriptions of postures, facial expressions, and side plays, and more pointed by the omission of material considered irrelevant. But at the same time, students of therapy who have come to find verbatim transcripts indispensable will find them – indispensable.

Though weak research-wise, the book is packed, 600 pages full, with discerning observations about the dynamics of interpersonal relationships in therapy and with good, practical suggestions for making group therapy work. At this stage of our knowledge of group therapy, the problem is not so much to validate its effectiveness, which the authors essayed with limited success, as it is to define the process and to describe rigorously just what goes on. This latter need the book meets admirably, for one approach to therapy and for a particular population. It will have much immediate value for therapists and will be a rich source of hypotheses for more circumscribed and definitive studies.

NICHOLAS HOBBS.

Peabody College.

GELLHORN, ERNST. *Physiological foundations of neurology and psychiatry.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. xiii +556. $8.50.

With the rediscovery of Freud and psychoanalysis during the thirties by organized American psychiatry, popular interest swung away from the organic etiology of mental disorder

and tended to concentrate itself upon "dynamic," developmental concepts and explanations. This break with tradition was a healthy one, but as is so often the unfortunate consequent of such extreme pendular movements within a discipline, there was not only an overemphasis upon functional factors, but a real neglect and almost a degradation of the physiological approach. It has seemed to this reviewer, however, that the last few years have seen the beginnings of a corrective movement with the pendulum swinging back toward interest in the physiological, and some hope for a more healthy balance between functional and organic influences in our study and interpretation of psychopathosis.

Professor Gellhorn's book is therefore a very timely one. Its explicit purpose is to survey and evaluate the links which can be established between experimental neurophysiology and clinical neurology and psychiatry. It begins with a survey of the basic factors regulating neuronal activity, proceeds through a discussion of the physiology and pathology of movement, the physiology of consciousness, and the physiology and pharmacology of the autonomic nervous system to a series of stimulating and provocative integrative chapters on neuro-endocrine action, the physiological basis of emotion, conditioning, homeostasis, and the phenomena of "constancy." The closing section on applications, relates the material to the clinical problems of the psychoses and psychoneuroses, and shock and carbon dioxide therapy.

This is admittedly a huge canvas, and no single author could be expected to do the picture complete justice. "Justice," however, is a relative term, and in this reviewer's opinion, Gellhorn has done an excellent job. There are some omissions, and it is obvious that Gellhorn writes with greater detail and greater enthusiasm where his own experimental work is concerned. He has not hesitated to take sides on controversial issues and occasionally he has made specific suggestions for treatment. This is illustrated by the chapter on the restitution of movement after central lesions where he pummels the general concept of the "plasticity" of neural function, and offers suggested lines to be followed in re-education. The net result of his partisan attitude, however, does not interfere with the fairness or scope of his survey, and does add a flavor and a definite liveliness to his treatment. Considering the difficulty of the subject, the book is well written and reads relatively easily. Gellhorn has succeeded in his goal of demonstrating "the fruitfulness of the physiological method for the study of pathological phenomena and the rewards for physiology itself of this type of research."

For psychologists the book is timely and important. In an age which is marked by its absorption in functional explanatory concepts, a devotion to higher order statistical abstractions, and a resulting tendency to intellectualize and verbalize the problems of human behavior, Gellhorn presents fundamental physiological data whose explanatory powers and predictive potential cannot be neglected. The book holds both a promise and a warning for contemporary psychology and cannot be overlooked.

WILLIAM A. HUNT.
Northwestern University.

SZONDI, L. *Experimental diagnostics of drives.* (Translated by Gertrud Aull.) New York: Grune and Stratton, 1952. Pp. x + 220. $13.50.

For two reasons the English trans-

lation of this volume, first published in German in 1946, is something of an anticlimax. Most psychologists in this country are now generally familiar with Szondi's test and unusual theories through the work of his student and co-worker, Susan Deri. Furthermore, research results on the Szondi Test appearing in recent years are generally unfavorable. There is little need to touch on the test itself in this review despite the fact that the book contains the basic manual and, presumably, the validating evidence for the test as a psychodiagnostic instrument. Borstelmann and W. Klopfer's excellent review and critical evaluation of the research on the test appeared in the March 1953 issue of this journal. Although Susan Deri claims that the test is not dependent upon the genetic theories of its author, Szondi apparently believes that data obtained from the test results verify his theories.

In *Experimental Diagnostics of Drives* Szondi presents a highly speculative theory of personality which, he states, is derived from a union of genetics and depth psychology (psychoanalysis) validated by the results of "more than 4000 experiments." The several thousand experiments evidently refer to that many individual administrations of the Szondi Test to an unreported number of subjects from an unspecified sample of the "general population," and to undescribed criterion groups of abnormal subjects. His preferred technique for citing evidence consists in anecdotal descriptions of single cases whose test profiles always seem to coincide exactly with the principle under discussion. The text also contains a number of percentage tables, but two of these are labeled "relative frequencies (assumed)," and the others are related in some undisclosed fashion to the "4000 experiments." Sample N's, means, and dispersions are not given.

This book is definitely not concerned with research on motivation, at least as we know it. The inclusion of the word "experimental" in the title and in many chapter headings is misleading. Thus the reader is forced to assess the personality theory on the basis of its reasonableness and internal consistency. To the American psychologist, the theory's reasonableness is immediately suspect from Szondi's preface which declaims that this "new approach serves as an independent means of psychodiagnosis in the service of psychopathology, vocational psychology, psychology of delinquency, pedagogy and characterology. . . . According to the working hypothesis repressed latent genes in the lineal (inherited) unconscious determine the choice in love, friendship, profession, sickness and death (*sic*)."

The role of dominant manifest genes in this psychology of predestination is not mentioned. Motivation, he claims, is based on eight specific drive needs (factors) derived as arbitrary polarities from four independent hereditary syndromes. He implies that the syndromes are accepted by geneticists interested in mental disorder. As there is no qualitative difference in motivation between the normal and the abnormal, the drive needs must be universals.

The book *may* contain some intuitive, basic insights into human personality, but, as far as can be judged, Szondi's insights are largely Freudian constructions translated into a new and less reasonable framework.

VICTOR C. RAIMY.
University of Colorado.

DVORINE, ISRAEL. *Dvorine pseudo-isochromatic plates.* (2nd Ed.) Baltimore: Waverly Press, 1953. Pp. 28. $12.00.

FLOYD, W. F., & WELFORD, A. T. *Symposium on fatigue.* London: H. K. Lewis, 1953. Pp. vii + 196. 24s

HARMS, ERNEST. *Essentials of abnormal child psychology.* New York: Julian Press, 1953. Pp. xiii + 235. $5.00.

HOVLAND, CARL I., JANIS, IRVING L., & KELLEY, HAROLD H. *Communication and persuasion; psychological studies of opinion change.* New Haven: Yale Univer. Press, 1953. Pp. xii + 315. $4.50.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. W., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953. Pp. xxii + 384. $6.00.

MALRIEU, PHILIPPE. *Les origines de la conscience du temps: les attitudes temporelles de l'enfant.* Paris: Presses Universitaires de France, 1953. Pp. 157.

NICHTENHAUSER, ADOLF, COLEMAN, MARIE L., & RUHE, DAVID. *Films in psychiatry, psychology and mental health.* New York: Health Education Council, 1953. Pp. 269. $6.00.

NOTCUTT, BERNARD. *The psychology of personality.* New York: Philosophical Library, 1953. Pp. 259. $4.75.

OSBORN, ALEX F. *Applied imagination; principles and procedures of creative thinking.* New York: Scribner's, 1953. Pp. xvi + 317. $3.75.

RHINE, JOSEPH BANKS. *New world of the mind.* New York: W. Sloane, 1953. Pp. xi + 339. $5.00

SARTRE, JEAN-PAUL. *Existential psychoanalysis.* New York: Philosophical Library, 1953. Pp. vi + 275. $4.75. (Trans. by Hazel E. Barnes.)

SCHEIFELE, MARIAN. *The gifted child in the regular classroom.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1953. Pp. x + 84. $.95.

SHAW, FRANKLIN J., & ORT, ROBERT S. *Personal adjustment in the American culture.* New York: Harper, 1953. Pp. ix + 388.

SPINELY, B. M. *The deprived and the privileged; personality development in English society.* New York: Grove Press, 1953. Pp. vii + 208. $4.00.

STERN, ALFRED. *Sartre; his philosophy and psychoanalysis.* New York: Liberal Arts Press, 1953. Pp. xxii + 223. $4.50.

STOLUROW, LAWRENCE M. (Ed) *Readings in learning.* New York: Prentice-Hall, 1953. Pp. viii + 555. $6.00.

VERNON, PHILIP E. *Personality tests and assessments.* London: Methuen, 1953. Pp. xi + 220. $4.00.

WEITZENHOFFER, ANDRÉ M. *Hypnotism; an objective study in suggestibility.* New York: Wiley, 1953. Pp. xvi + 380. $6.00.

Psychological Bulletin

EFFECTS OF EARLY EXPERIENCE UPON THE BEHAVIOR OF ANIMALS

FRANK A. BEACH AND JULIAN JAYNES
Yale University

It has always been apparent that human beings learn a great deal during infancy and childhood and that the results of this learning often continue to affect behavior throughout the remainder of the life span. Long before the days of a science of behavior, naturalists speculated about the possible importance of early learning in the lives of lower animals. More than two thousand years ago Aristotle wrote:

Of little birds, some sing a different note from the parent birds, if they have been removed from the nest and have heard other birds singing; and a mother nightingale has been observed to give lessons in singing to a young bird, from which spectacle we might obviously infer that the song of the bird was not congenital with mere voice, but was something capable of modification and improvement (5, p. 121).

During the two thousand years after Aristotle other writers compiled anecdotes and life histories of various animal species, but not until the emergence of comparative psychology in the closing years of the nineteenth century did the subject of behavioral development attract serious scientific interest. Among the earliest American workers to deal with this problem was Small, who in 1899 published an article on the albino rat as a useful subject of psychological investigations (119). Small's account included a section dealing with behavior in infancy and adolescence. Eight years

later, Yerkes' book, *The Dancing Mouse* (132) appeared, and that treatise dealt among other things with ontogenetic aspects of behavior.

In recent years the amount of attention paid to this subject has increased significantly, and the change is attributable to several different sources of stimulation. Some of the first systematic experiments dealing with behavioral development were generated by the problem of the relative importance of "maturation" and "practice" in the perfection of simple response patterns, such as the pecking behavior of young chicks (117) or the swimming behavior of larval amphibians (16, 47). Additional impetus was given to investigations in this field by the impact of Freudian theory (38), which led to various studies, such as those which limited food supply or feeding responses in young animals and then measured the effects upon certain types of behavior in adulthood. A third influence, slower to gain recognition in America, has derived from the observations and theories of Konrad Lorenz and other European biologists who examined the effects of early social stimulation upon the adult behavior of various species of birds.

The most recent increase of interest in the effects of early life experiences upon the behavior of adult animals is traceable to theories that

stress the importance of perceptual learning in infancy upon subsequent performance in tests of learning. A leader in this field is Hebb, whose *The Organization of Behavior* 52 has been directly or indirectly responsible for a number of experiments reported in psychological journals during the last two or three years.

Because of the rapidly growing interest in this area of comparative psychology, we have assembled and evaluated as much as possible of the relevant evidence. This seems worth doing not only because of its usefulness to students of animal behavior, but also because scientists interested primarily or exclusively in human psychology very frequently turn to the literature dealing with lower animals in connection with discussions of human development. The area covered is a diverse one; it touches and illuminates many fields of scientific investigation; its findings bear on a wide variety of problems ranging from plant parasitism to human personality; and its significance stretches from the genesis of sensory experience to the theory of evolution. Since the resulting diversity is very extensive, we have found it most feasible to organize our review in terms of the types of behavior which investigators have measured after modifying the animal's early experiences.

As the reader will discover, we have given the term "experience" a very broad meaning and used it to include a variety of factors ranging from the type of food consumed by larval insects to the characteristics of visual stimulation experienced by infant chimpanzees. This plan has been followed because of our conviction that there is no clear-cut line of division between the effects of early "psychological" experience and certain physiological antecedents of behavior. The development of behavior can be examined from either the psychological or the physiological viewpoint, but this is an arbitrary, methodological distinction made possible for the sake of convenience. Any analysis which is confined to one of these alternatives is bound to be incomplete and to a greater or lesser degree misleading.

SENSORY DISCRIMINATION AND PERCEPTION

Descriptions of the sensory and perceptual abilities of animals have very often implied the unwritten suggestion that these abilities are exclusively determined by innately organized neural mechanisms. Several recent studies tested this assumption by studying the effect of early sensory experience upon the sensory abilities of the adult.

Vision. Visually directed responses of laboratory rats are not modified to any marked degree if the animals are prevented from using their eyes during infancy. Animals of this species which have been reared in darkness proportionally adjust the force of jumping to a food platform at different distances, although during the first trials the accuracy of such jumps is inferior to that of normally reared animals (75). Rats reared in darkness are capable of size, brightness, and pattern discrimination after very little visual experience (49, 50), and correspond to normal animals in their ability to generalize from one visual stimulus to another on the basis of relative size or relative brightness (50).

Investigations of the importance of early visual stimulation in other subprimate vertebrates have shown more pronounced effects, and suggest the existence of marked differences

between species. It is reported, for example, that when a rabbit which had been reared in darkness was first brought into the light, it assumed abnormal bodily postures and bumped into obstacles as it moved about the environment (40). The eyeblink reflex in the dark-reared rabbit was normal, but the latency of pupillary response to light was retarded; it was five seconds in duration. After one month under normal illumination the visually controlled movements of this animal had improved to normal.

This result was not attributed to anatomical degeneration, since examination of the retinas and optic tracts of other rabbits reared in darkness revealed no structural abnormalities. Recent work, however, contradicts this earlier conclusion. Chemical analysis of the ganglion cell layers of retinas from rabbits reared for the first ten weeks in total darkness showed a complete absence of the pentose nucleoproteins present in normally reared animals (13). The cause is presumed to be the lack of adequate visual stimulation during the postnatal period of development. It is therefore possible that the visual deficit found in dark-reared animals is attributable to incomplete chemical development of the ganglion cells in the retina.

Pigeons whose eyelids had been sewn together before they had opened behaved abnormally when the lids were opened six weeks later (88). Like the rabbit reared in darkness, these birds tended to assume abnormal postures and apparently were unable to avoid obstacles on the basis of visual cues. Furthermore, they did not show the usual startle response when a hand or handkerchief was passed rapidly before their eyes. Some reflexes, including pupillary responses to light, were unaf-fected by the lack of previous visual stimulation, but optokinetic nystagmus was impaired. In general, posture, obstacle avoidance, and visual responses improved rapidly and were normal three days after the lids had been freed.

A similar, partial visual deprivation was produced by another experimenter (118), who encased the heads of ringdoves with translucent vinyl hoods when they were three days old, before their eyes had opened. This technique permitted diffuse uniform light to reach the retina, but prevented form perception. Two months later, some opacity was found in the vitreous humor of the eyes, but contrary to the previously cited study, there were no differences in the optokinetic reactions of the hood-reared doves and normally reared birds. No observations on obstacle avoidance or posture are reported, but when first unhooded and forced to fly, the ringdoves did so erratically and with definite errors in space orientation. For purposes of some future experiments, the experimenter cut windows in the hoods of the birds when they were two months old, so that only the lower half of the eye could be used, and then tested the birds on a form-discrimination problem. It was found that the hood-reared birds took significantly longer to learn a circle versus triangle discrimination than similarly tested controls, the conclusion being that early experience is a prerequisite to normal adult function in visual perception.

The deleterious effects of prolonged visual deprivation are not limited to mammals and birds. Eyed fishes taken into light from lightless subterranean cave pools repeatedly bump into obstacles when swimming about in a strange aquarium. They fail to

seize food as it drifts downward through the water, but instead seem to rely on chemical sensations as they grub about in the sand covering the tank floor. Visual control of swimming and feeding appears gradually and is only complete after several weeks under normal illumination. Fish that were reared in the laboratory in lightproof tanks for two and one-half years behaved like those captured in underground lakes and did not show normal visual responses until they had spent about one month in the light (14).

Most primates depend heavily upon visual cues, and it is not surprising that lack of visual stimulation in early life has pronounced effects upon subsequent behavior. Chimpanzees that were kept in the dark for the first 16 months of life were comparable to the pigeons described above in that they showed a marked pupillary response to increase in light intensity, but optokinetic reflexes were impaired and there was no blink when an object suddenly approached the eye (105). These apes exhibited head and eye movement to follow a point of light in a darkened room, but steady fixation and obstacle avoidance on the basis of visual cues were totally absent. In one case approximately 50 hours of visual experience elapsed before the first signs of visual learning (reaching for the nursing bottle) appeared (105). These results may have been due to physiological changes resulting from a lack of visual stimulation, since ophthalmoscopic examinations at a later date revealed evidence of severe retinal and optic nerve degeneration (107).

In a second experiment using chimpanzees, one animal was reared in total darkness to the age of seven months. A second was reared in the dark but exposed to diffuse light (no image formation possible) for one and one-half hours per day. A third ape was taken from the dark each day and allowed to spend one and one-half hours in a normally illuminated room (106). Tests begun when the animals were seven months old showed that the individual with full visual experience for 90 minutes per day behaved like a normal chimpanzee. In the partially deprived ape (exposed only to diffuse light), the blink response to objects approaching the eye appeared after six days of visual experience, while in the totally deprived ape (no light during the first part of the experiment), the blink reaction did not appear until 15 days. Steady fixation appeared in 13 days in the partially deprived, and 30 days in the totally deprived animal. Learning to avoid a conspicuous visual stimulus which was associated with electric shock took one or two trials in one day for normal apes, 13 days of two trials per day for the partially deprived individual, and 15 days of two trials per day for the chimpanzee reared in total darkness. Ophthalmoscopic examination revealed evidence of retinal degeneration in the totally deprived animal, but no ocular abnormalities were found in the other two animals (107).

The evidence is fragmentary and at points inconclusive, but it does suggest that in at least some species of fishes, birds, lower mammals, and primates absence of the normal amount of visual stimulation during the developmental period may result in inability to respond adaptively to visual cues when such cues first become available to the individual. The defects do not appear to be permanent but seem to disappear with in-

creasing experience in visually directed responses.

Touch and proprioception. There is some reason to believe that some other types of sensory deprivation in infancy may affect subsequent behavior. One anecdotal account describes a young chimpanzee whose paws were badly injured in infancy and which, as a consequence, never indulged in the frequent climbing behavior typical of this species. It is reported that after its paws had healed, this animal rarely exhibited any tendency to climb and showed signs of dizziness whenever it got more than a few feet above the ground (53).

A more recent experiment involved bandaging or encasing an infant chimpanzee's lower arms and legs in cardboard tubes and thus preventing the normal manipulative use of hands and feet (92). The restraining devices were in place, on the single subject, for 30 months, beginning with the fourth week of life. When the limbs were freed, the ape showed some forced grasping, which normally disappears during the first postnatal month. Walking and sitting postures were abnormal. Grooming behavior, so typical of the normal chimpanzee, was lacking, as was the almost universal tendency to cling to the human attendant when picked up.

When a normal chimpanzee is touched lightly at some point on the body, the usual response consists of promptly bringing one or both hands to the locus of stimulation. In the experimental animal these responses were either lacking or abnormally slow and inaccurate. Though visual discrimination in this animal was normal, the ape had great difficulty in learning to differentiate between tactile stimulation of the right and left hands when vision was excluded.

FEEDING BEHAVIOR

Pecking responses in birds. As noted in our introduction, the controversy regarding the relative importance of maturation and experience in the shaping of behavior led to several studies of pecking behavior in chicks (12, 25, 87, 95, 112). These have been reviewed so often that they need not all be discussed here (see 80). In general, it appears that birds which are prevented from pecking by being reared in darkness, and then tested for the accuracy of such behavior, tend to be somewhat inferior to normal controls at first, but to improve rapidly and equal fairly soon the performance of chicks that have not been deprived. The most interesting report from the point of view of the present discussion is the statement that dark-reared chicks, fed artificially from a spoon for two weeks after hatching and then placed in the light, never developed the pecking response. Put in a pen with an ample supply of grains and grit to peck at, "they all starved to death even in the midst of plenty" (95, p. 436).

As noted earlier in this review, some animals that have been reared in darkness are incapable of normal visuomotor behavior. It may be that part of the failure of dark-reared chicks to peck normally is due to sensory or perceptual deficiencies rather than to delayed motor development alone. Preventing the pecking response for long periods after birth has not been found to cause the disappearance of that response in other birds. Ringdoves, for example, were artificially fed from three days of age and prevented from pecking by being reared with translucent hoods over their heads (118). After

as long as two months, the hoods were removed, and the birds developed coordinated pecking at grain within 15 hours.

The visual stimuli capable of evoking the pecking response can be modified by early feeding experience. One group of chicks was reared for seven weeks in a cage which was illuminated exclusively from above, while in the cage of a second group the only light came through the translucent floor. Later the response to photographs of grains of corn was tested. In some pictures the corn was illuminated from above, in others from below. The chickens reacted selectively, pecking only at those photographs in which the lighting conditions corresponded to those under which real feeding had taken place (55).

A different approach to problems of pecking activity has been suggested by Levy's theory that there exists a pecking "need" or "instinct," and that birds continue to peck until this need is satisfied or satiated. To test this hypothesis, one large group of hens was raised in a cage having an elevated floor of wire mesh, while a control group was kept on a dirt floor (77). Chickens kept on wire pecked at their cage mates until many birds were partially denuded of feathers. Comparable behavior was rare among the control birds, which could peck the ground. When transferred to ground cages at six weeks of age, and mixed with control cases, the chickens reared on wire vigorously pecked at the backs of the fully feathered controls, but this behavior died out after two months. Another group of chicks was reared on wire floors for a longer period of time. It was found that the pecking of companions was progressively reduced and ceased after four months, but when transferred to the ground

at seven months of age these birds again began to peck one another as vigorously as before.

Nursing behavior in mammals. Levy has suggested a comparable theory for the sucking responses of mammals. It is proposed that under normal conditions, a "need" to suck is satisfied as the infant sucks the mother's teat or breast, but if the need is not satiated during nursing, nonnutritional sucking behavior will appear. The occurrence of chronic thumb sucking in some bottle-fed babies and the reported ear licking and tail sucking of bucket-fed calves have been interpreted in this way (76). In an attempt to test the hypothesis, four puppies were separated from the bitch at 10 days of age and fed from nursing bottles thereafter (76). Two dogs drank from nipples with very small holes so that to obtain their milk ration they sucked an average of 80 minutes a day. The other two animals were fed from large-holed nipples and consumed the milk supply in an average of 13 minutes of nursing per day.

It is reported that the rapidly fed puppies exhibited sucking movements after feeding time and made sucking noises while asleep. In addition these dogs tended to suck their own bodies and those of other animals as well as to lick inanimate objects in the kennel. None of these characteristics appeared in the two animals that had the longer periods of nursing during infancy. A more recent study has investigated the possibility of producing this effect through sucking deprivation at an earlier age (109). Three puppies were partially deprived of sucking experience but adequately fed for the first 10 days of life and then returned to the mother. During the deprivation period the experimental animals showed a stronger

sucking pressure and more nonnutritional sucking than normal controls. These differences disappeared, however, when the dogs were returned to the bitch, and there was no persistence of the compensatory sucking such as had been reported in the earlier experiment.

Hoarding behavior in mammals. Laboratory rats do not normally hoard their food, as do many other rodents, but they can be induced to do so under appropriate conditions. The evidence is somewhat contradictory, but the results of some investigations indicate that early experience with food may have some effect upon hoarding in adulthood. According to the first study of this problem, rats reared on powdered food hoard fewer food pellets than other animals that have always eaten pellets; and subjecting rats to semistarvation in prepuberal life increases their tendency to hoard food in adulthood, even though the animals are satiated at the time of testing (131). Results of a subsequent experiment suggested that partial starvation for 15 days immediately after weaning produces increased hoarding in adulthood, but only if the hoarding tests are preceded by a period of food deprivation (59). In this same investigation it was found that partial starvation later in life, beginning 12 days after weaning, has no effect on food hoarding in adulthood. Attempts to reproduce these results in later experiments showed hoarding differences between rats starved in infancy and controls, but the differences were not as statistically reliable as those found previously (60).

A more recent attack upon this general problem involved study of the effects of infantile deprivation on the hoarding of water-soaked cotton as well as food (79). Albino rats of both

sexes were divided into three groups. For 15 days after weaning, one group was partially deprived of food, and a second group of water. A control group was maintained on ad libitum food and water until adulthood. At 130 days of age, all animals were placed in special cages where hoarding tests were conducted, both before and after a period of deprivation. The food-deprived males were found to hoard significantly more food pellets than controls after adult deprivation, but no significant differences between deprived and control animals were found among the females, or in the water-hoarding tests for either sex. The picture presented by such results is of a more complicated problem than was at first supposed.

Positive results in studies of this type may be due to the prior acquisition of pellet-seizing and pellet-carrying responses, as well as to an increased rate of eating in the experimental animals (82). This hypothesis has been tested, and it was found that rats deprived severely in infancy had a faster eating rate as adults (83). Presumably, such rats would need less time to eat in the hoarding tests and would therefore have more time to hoard extra pellets.

The positive reports of increased hoarding consequent to infant feeding frustration have been rather widely cited as bearing upon psychiatric theory (57, 104). In view of the lack of clear-cut confirmation by subsequent experimenters, final judgment probably should be reserved.

REPRODUCTION

Among the most species-specific behavior patterns in any particular species are those having to do with reproduction—courtship, fertilization, nest building, parturition, and the

care of the young. For a long time, these often unique stimulus-response organizations were classified as "instincts" and were thought to be exclusively determined by genetic constitution. Evidence which shows that the occurrence of these responses is heavily dependent upon other factors, especially the early experience of the animal, has recently been gathered from many widely divergent invertebrate and vertebrate species. It is worth stressing at this point that early experience appears to influence the kinds of stimuli that will later evoke reproductive activities, whereas the actual responses are less subject to modification.

Egg-laying behavior in insects. Several studies have shown that the particular chemical stimuli which call forth the egg-laying responses of gravid females of several insect species are determined in part by the type of food which the insect has eaten or been exposed to during the larval stage of life. The female sawfly normally deposits her eggs on the leaf of a particular species of willow tree, but if forced to do so she will oviposit on a different kind of willow. In an early experiment, successive generations of females were forced to deposit their eggs on an unpreferred species, and, after four years, the descendents showed a preference for the new type of host plant (**46**). However, the mortality in this study was high, and the results could possibly be explained on the basis of genetic selection.

More carefully controlled studies have employed a choice method in which the insect's tendency to approach different hosts is tested in a Y-shaped tube. Thorpe and Jones (**128**) found that all adult females of the ichneumon fly are attracted to the odor of meal-moth larvae, which are

the normal hosts. But insects that had been reared on wax-moth larvae during their own larval period were also attracted to the odor of larval wax moths, while insects reared on normal hosts were not. Immediately after they had hatched, normally reared females were brought into contact with meal-moth larvae, and as a result of this exposure they showed temporary attraction to the immature meal moths; but the attraction was lost if the ichneumon females were isolated for 10 days (**123**). Similar changes in adult preferences produced by varying the type of food consumed in the larval stage have been reported for fruit flies (**124**) and vinegar flies (**26**).

Thorpe (**124**) regards this type of modification of behavior in insects as a conditioned response and comparable to imprinting phenomena in birds, which will be discussed later in this review. Whatever the explanation, these experiments do demonstrate: (*a*) that adult olfactory preferences in some invertebrates can be affected by early environment, and (*b*) that this type of effect can in turn influence egg-laying behavior in these particular species. It is important to add that this conditioning effect does not occur in all insects. When females of the Colorado beetle were reared as larvae on a foreign type of host plant, they did not oviposit on the new plant (McIndoo, as cited in Thorpe, **124**), and descendents of a strain of chalcid wasps which had been confined exclusively to a foreign host for 260 generations still preferred the ancestral host (**111**).

Sexual selection in fish and birds. The reproductive behavior of some fishes and of many birds is characterized by the comparative importance of elaborate patterns of display and courtship which are usually an indis-

preamble to the mating act itself. The overt responses involved do not appear to be modified by prior experience, but in some cases the stimuli that elicit display and courtship have been found to be markedly dependent upon certain kinds of early experience.

Although adult male cichlid fishes apparently show courtship and mating responses only toward members of the opposite sex, males that were initially reared in isolation attempted to mate indiscriminately with both males and females (93).

Passenger pigeons reared by foster ringdove parents are said to have failed to mate with their own species as adults, but to have attempted sexual union only with birds of the same species as their foster parents (23). Other ornithologists have noted similar interspecies "sexual fixations." Turtledoves, domestic and wood pigeons, and magpies have been reported to pair with individuals of other species if they are reared with that species (41). Unfortunately, the lack of experimental controls renders the interpretation of such observations difficult, but the results suggest that in some species of birds, individuals tend to mate with the species with which they have been associated in infancy.

Comparable effects upon adult mating preferences are sometimes found in birds that have been reared by human beings. Three male, blond ringdoves were reared singly by Craig (24). When they were first put into a cage with a female dove, these males directed their courtship and coital responses to the hand of their caretaker or to inanimate objects in the cage. Eventually, normal mating between the birds occurred. It has been reported that a domestic turkey cock, which had been reared by hu-man males, responded to men with the usual courtship and mating activities, whereas women elicited the fighting and retreating reactions normally reserved for rivals of the same sex (101). This cock never displayed to turkey hens, but would tread them or any similarly sized object if he were forcibly placed in the mating position.

Mating behavior in mammals. No interspecies mating preferences due to special conditions of rearing have been reported for mammals, but other effects of early experience on sexual activity have been described. One early study of male rats led to the conclusion that prepuberal segregation of males increases the tendency of an animal to cross an electric grid when a male is on the opposite side and decreases the tendency to cross to a female (65). However, the interpretation of results obtained with the obstruction method is difficult, and the unsystematic procedure of this experiment places its results in serious doubt. Other studies (7) do not support the suggestion that preference for a masculine sex partner is influenced by the conditions of rearing. Rearing rats in isolation has no effect on the pattern of the copulatory response itself either in males (7) or in females (122), but it does tend to increase the proportion of males that copulate in comparison with other groups that have been reared with females or with other males (7).

An effect of prepuberal experience on adult sexual activity in the male rat has recently been reported (66). Male rats were reared in individual cages from infancy, but periodically throughout the prepuberal life phase they were exposed to age mates of either sex in standardized ten-minute tests. When tested with sexually receptive females in adulthood, these

males displayed some copulatory responses, but also indulged in a great deal of the wrestling and play behavior that is characteristic of the immature rat.   Ejaculation in the male rat normally follows a series of closely spaced intromissions, but in the experimental males, the frequency of intromission was greatly reduced by the interpolation of play reactions, and ejaculatory responses were therefore infrequent.

Parental responses in mammals. Care of the young depends in some instances on experience gained before sexual maturity. Early experiments revealed that primiparous rats, confined in individual cages up to the time of parturition, made use of available nesting material and cleaned and cared for their young in a biologically adequate fashion although they had never seen nests built or delivered litters (6). It appears, however, that nest building is in some fashion related to earlier experience in the manipulation of objects in the environment. Female rats that were reared in cages containing nothing which could be picked up or transported are reported to have built no nests at parturition, even though the appropriate type of materials was available (108).

In an as yet unpublished study, female rats were reared with rubber "Elizabethan ruffs" around their necks (Birch as cited in Riess, 108). This technique prevented the female from washing herself or from having any mouth contact with her own genitalia.   The ruffs were removed shortly before parturition, and it is reported that the experimental females failed to clean or to nurse their young.

Effects of captivity or domestication upon reproductive behavior. The effects of rearing animals in captivity sometimes include the appearance of sexual maturity at an earlier age than

has been observed under natural conditions, as well as extension of the breeding periods in adulthood. This occurs in a variety of mammalian species including the ibex, wild boar, American bison, antelope, chamois, and others (53). In one controlled study, the offspring of wild grey rats (*Norwegicus*) were bred for 25 generations in the laboratory. In the course of this time, the length of fertility in relation to the life span was doubled, although average litter size did not change (71). Some mammals, however, such as the polar bear (120), show no alteration in the length of the breeding season under captivity, and failure of normal sexual development has been reported in three gorillas reared in captivity since infancy, and also in the cheetah and Malayan bear (53).   Whether these effects have some physical cause, dietary or otherwise, or are due to early experience is not known.

GREGARIOUS AND FILIAL BEHAVIOR

Some species of animals maintain colonies, schools, flocks, or herds throughout the year.   In other species, formation of social groups is limited to the breeding season or to the time during which the mother and young must remain in close association. In both instances, study of the influence of early experience on social behavior has yielded important results.

Species recognition in insects and fish. Among invertebrates, such aggregations as have been studied seem to depend on some single specific stimulus. For example, each colony of ants, even within the same species, has a unique odor. Members of a colony recognize each other by means of this cue, and an ant contaminated with the odor of a different colony is usually attacked and killed (9). It is

possible that this olfactory basis of social organization is partly determined by early experience, as is host preference in some parasitic insects. If normally antagonistic species of ants are placed together within 12 hours after emerging from the pupae, there is no aggressive behavior between them, and a permanent mixed colony can be established (32). The effects of this early "conditioning" are apparently permanent, for temporary separation of a mixed colony of two species after a month of association did not decrease their compatibility. Even after seven months of separation, ants that were originally reared in a mixed colony did not attack those of the same odor with which they had been reared, though all other "foreign" ants were promptly attacked and killed (33).

Visual stimuli are commonly involved in the social interaction of several kinds of fishes, and in some species responsiveness to a given visual stimulus is influenced by early experience. Adult jewel fish of one type show a blue-black color of the ventral surface during the brooding period. The fry of this species normally have a positive response to blue-black objects and are unresponsive to red. But young fish that have been reared by foster parents with brilliant red underbellies show a reduced response to blue-black, and a positive reaction to red (94). In other words, the sensory basis for filial grouping in these fishes is partly dependent on very early experience.

Gregariousness in insects. The color, morphology, and aggregation behavior of solitary and migratory locusts differ quite markedly, and at one time they were classified as separate species. It is now known that these striking differences are "phases" of the same species (129), and that the differentiating charac-

teristics are directly related to certain conditions existing during the insects' development. In an early experiment it was found that progeny of the solitary locust could be transformed into the migratory type by rearing them under crowded conditions, whereas individuals reared in isolation never become migratory (31). This relationship of early environmental conditions and adult behavior in the locust has been confirmed in field observations (70, 130).

A well-designed experimental analysis of social behavior in locusts has recently been performed and further strengthens the above conclusions (28). Groups of locusts in the first instar (i.e., from hatching to the first metamorphosis) were reared under crowded conditions, and others were reared in individual jars. When the experimental animals had metamorphosed, and were therefore in the second instar, aggregative tendencies were measured by placing groups of them in a ring-shaped cage, the floor of which was marked off in equal-size segments. At intervals after the beginning of the test, the numbers of insects in each cage segment were counted. By comparing the frequency distribution of segments containing various numbers of individuals with a Poisson series it was determined that locusts which had been reared in groups aggregated at a level well above chance, whereas insects reared in isolation were significantly less gregarious.

When the formerly isolated locusts were kept under crowded conditions for 24 hours, there was a considerable increase in gregarious behavior as measured by the same test, and after four days of group living, their social behavior was identical with animals reared in groups since hatching. In contrast, the social behavior of locusts reared for the first instar

in crowded conditions was not appreciably affected by subsequent isolation. This suggests that in this species the tendency to aggregate, once it has been acquired in infancy, is irreversible.

Gregariousness in birds. The effects of early experience on social behavior seem to be much more complex and extensive in most birds than in fishes. Lloyd Morgan observed that domestic chicks reared in an incubator are relatively indifferent to the hen and her flock (86), and recent experiments indicate that hand-reared chicks respond very slowly, if at all, to the food-call clucking of the hen (20). Mallard and Muscovy ducklings reared in an incubator do not show the normal selective response to adults of their own species. Qualitative measures of the tendency of isolated chicks to move toward another chick in a runway revealed that the approach was much slower than normal, with frequent pecking at the second bird. When they were placed with the flock, such chicks tended to keep apart from the group (19). These antisocial effects are maximized by rearing chicks in complete isolation away from any animal or human being. Chicks so reared are said to have developed a persistent and "rage-like" restlessness as well as unusual stereotyped behavior patterns, and to have reacted in a "fearful" fashion when placed with the flock (Bruckner, as cited in Katz, 68).

There is some reason to doubt that rearing chicks in isolation completely abolishes all positive responses to others of their kind. Chicks have been reared in isolation to the fourth day after hatching, and then tested for their choice between a stimulus compartment containing chicks and one containing mice. They spent more time in front of the former (97).

Interestingly enough, birds that had been raised in the flock showed no such preference.

"Imprinting" in birds. In studying reproductive behavior of birds, we noted several instances in which the choice of sex object by the adult bird was influenced by the species with which the individual had been associated in infancy. That other social responses can become "fixated" in this way, especially toward human beings, has long been known. It is said to be true, for example, of the rhea bird (58). If eggs of the European greylag goose are hatched in an incubator and the young are exposed to a human being just after hatching, the young birds respond in filial fashion to the human but not to adults of their own species (Heinroth as cited in Lorenz, 78). It is further claimed that incubator-hatched geese can be reared by mother geese only if they are placed in a sack immediately upon removal from the incubator, so that premature exposure to humans is prevented. To describe the process responsible for these phenomena, the German ornithologist, Konrad Lorenz, has coined the term *Prägung*, translated as "imprinting." Imprinting takes place when the first stimulus to evoke a given "instinctive" reaction becomes, with extreme rapidity, the only stimulus that henceforth can evoke that reaction. According to Lorenz, imprinting has two properties which distinguish it from other types of learning. First, it can only take place during a specific and brief portion of the life span while the organism is in a critical physiological stage of development; and second, the results of imprinting are irreversible throughout the life of the individual.

Lorenz confirmed Heinroth's observations on the greylag goose. Newly hatched goslings were exposed

to and allowed to follow the experimenter and were prevented from seeing, hearing, or being in contact with adult geese until the tendency to follow the human foster parent had been firmly established. Thereafter, they tended to follow and remain close to Lorenz and showed no positive social responses to adult geese (78). There is evidence to show that imprinting occurs in some other birds, especially in those ground-nesting species whose young are capable of locomotion and self-feeding as soon as they are hatched. Young chickens, ducks, or turkeys have been reared by foster parents belonging to one of the two other species and under such conditions have been found to form cross-species filial attachments (102).

A number of independent studies combine to indicate that susceptibility to imprinting is a function of the age of the young bird. The coot is most easily affected by imprinting during the first 8 hours after hatching, and coots that are 24 hours old cannot be imprinted (1). The most thorough study of filial imprinting has been made on ducklings (29). The young birds (tufted ducks, eiders, and shovellers) were hatched in an incubator under opaque hoods. At various durations after hatching, the hoods were removed, and the ducklings were allowed to follow the retreating experimenter. It was found that for all three species the following-reaction could be evoked, and imprinting established most easily in the first 6 hours and, as in the coot, could not be obtained if 24 hours had elapsed since hatching.

Not all birds followed the experimenter without his also making auditory signals. Such sounds increased the probability of evoking the following-reaction and at least in the tufted duck, the shoveller, and the mallard seemed to have a stronger releasing

effect than visual stimuli. Birds that had been imprinted subsequently followed the experimenter whether he walked, swam, or rowed a boat, and paid no attention to adult ducks. After three weeks, the following-reaction to humans gradually diminished.

In tree-nesting birds, whose young have a slower behavioral development, the formation of social attachments takes a somewhat different course. The initial phase of imprinting to a parental object is believed to occur as it does in ground-nesting birds, but the results do not become irreversible until some time later. For example, jackdaws taken from the nest at 14 days of age were found to be imprinted on their natural parents, but after separation from adult jackdaws and continued association with humans from this time on, the birds' positive social responses could be shifted to human beings (78). After 20 days, when the young jackdaws are fledged, such shifts in social responsiveness cannot occur. Those that have been with adult birds cannot be imprinted to humans, and those that have earlier been imprinted to humans cannot be induced to react normally to their own kind. These preliminary findings on one insessorial species are of interest, but they require a more thorough conceptual and experimental analysis before they can be generally accepted.

There are some avian species in which parental imprinting to humans does not occur. For example, some wading birds, such as curlews and great godwits, show the escape reaction to man immediately after hatching in an incubator, and hence imprinting on man is impossible (78).

It is interesting to consider the possible role of imprinting in the evolution of species. It seems highly probable that it operates as a social

and reproductive isolating mechanism in the case of sympatric species. That it may also play a role in fostering the geographical isolation of some allopatric species should be of interest to ecologists. Thorpe has pointed out that the significance of imprinting may not be confined to the formation of bonds with other living creatures, but may apply also to bonds with inanimate aspects of the environment (126). Evidence for this is scant, but it is known that acoustic signals (29) or boxes and footballs (102, 127) can be used as imprintable stimuli for various birds. Validation of this hypothesis might aid in understanding the formation of ecological races. For example, peregrine falcons nest in cliffs in one region and in trees in another, but in periods of overpopulation, the cliff-district birds unadaptively attempt to nest in unsuitable cliffs rather than switch to trees like falcons of other districts (56). This may be due to imprinting of the young falcons to cliff nests.

Song acquisition in birds. The songs of birds are an important mechanism in their social organization, especially in the breeding season, and it is widely believed that in some cases these songs are influenced by auditory experience acquired early in life. For some species, such as the rhea bird (100) and roller canary (85) no such influence is found, but if young common canaries are allowed to hear the recorded song of the nightingale they develop a song intermediate between the normal canary and nightingale songs (61). In another study, bluebirds, robins, woodthrushes, orioles, bluejays, and redwinged blackbirds were reared together in a large room. At maturity, none of these birds sang the normal song of its species (115). Similar modifications of the adult song resulting from conditions of rearing have

been reported for other species including bobolinks, grosbeaks (116), sparrows (21), whitethroats, tree pipits, meadow pipits, greenfinches, chaffinches, and yellow buntings (54).

Song acquisition in birds has been considered as a special type of imprinting (127). This interpretation is supported by the report that some birds, like the chaffinch, never acquire the normal song if they have developed an abnormal one in early life (43). That this also may be a case of an isolating mechanism directly related to early experience has been suggested by Mayr (84) in his discussion of the song races of chaffinches about Stuttgart. A male chaffinch, who had by some chance association with other birds learned a different song in infancy, would be expected to employ this song in courtship. If he were successfully mated, the progeny would then be exposed to the special song, acquire it, and use it in their courtship later, and in this manner a new species or subspecies might develop. Thus it is seen that in birds early experience can exert a marked influence upon the future social life of the individual, and also perhaps upon the future of the species.

Social behavior in mammals. Studies of the social effects of early experience in mammals are relatively few. Naturalistic observations reveal that lambs of wild sheep show something like the imprinting process in birds. When captured and reared by humans, they do not later rejoin the herd, but follow and remain with human beings (89). Domestic lambs that have been reared on a bottle exhibit similar behavior. When turned to pasture with the flock, they show difficulties in adjusting to other sheep and tend to graze by themselves (112).

Scott and Marston (113, 114) have

outlined a theory to the effect that the development of young puppies is divisible into several sequential periods, certain of which are "critical" in the sense that during these relatively brief phases the young dog forms habits or develops tendencies which have profound effects upon social behavior in later life.

There is no experimental evidence to establish the point, but it may be noted that the parturitional behavior of certain wild ungulates is such as to allow maximal opportunity for newborn young to form social bonds to the mother within a very short time after birth. Female red deer and American elk associate in herds throughout most of the year, but just before calving, the pregnant female leaves the group and seeks seclusion. For several days after parturition the young remain isolated and hidden, and are visited periodically by the mother. Then both the female and her offspring rejoin the herd, and thereafter the calf's following and suckling responses are directed solely to its own mother, while she in turn permits only her own young to remain close to her and to nurse (2, 27). The possibility suggests itself that during the brief period when the calf's social contact is restricted to its mother, imprinting or some similar process takes place.

The chimpanzee is a highly social animal and may be considerably disturbed by temporary isolation (73, 91). The object of its social impulses appears to be influenced by experience in infancy, just as it is in many birds and in some subprimate mammals. Several observers (48, 63, 69) who have reared infant chimpanzees in their homes have noted how easily at this age the social reactions of these animals can be transferred to human adults. One such chimpanzee was later returned to the laboratory and caged with others of its own species. For several months he did not mix with other apes, but looked "longingly" at human beings who were visible to him beyond the cage (Finch, as cited in Nissen, 91).

EMOTION AND TEMPERAMENT

Wildness and timidity. To the human observer, animals in nature display characteristic behavioral tendencies that are often described as wildness, timidity, or aggressiveness. Some of the laboratory studies cited in this review indicate that certain emotional responses in animals are often directly related to experiences acquired early in life. In birds, we have seen how rearing in isolation produces unusual emotional behavior (Bruchner, as cited in Katz, 68), fearfulness when placed with the flock, and a tendency to peck at other birds (19). Another study involved rearing domestic chicks in small wooden boxes inside the brooder where other chicks could be heard and their shadows seen. Even this partial isolation resulted in fighting behavior when each of these birds was paired with a normal chick at seven weeks of age. Fear reactions and "bewilderment" when placed with the flock were also reported (15).

"Wildness" involves both genetic and experiential factors. Among the latter, one significant natural variable for some species seems to be the temperament of the mother. Black ducks and mallard ducks are equally wild in nature, but the mallard's wildness is readily modifiable when the birds are hatched and reared by tame foster parents (98). Wild herring gulls that have been reared by the somewhat tamer common tern do not display the usual herring gull's timidity toward man (Kuhleman, as cited in Thorpe, 127).

The tameness of the mother seems

to be an important factor in the wildness of some small mammals. Rats reared by tame females are more docile toward humans than those reared by wild mothers (103), though the taming effect is said to be cancelled out if a wild father remains in the same cage with the family (110). In another study it was found that descendents of wild gray rats reared in captivity for 25 generations lost their "high nervous tension and fear of man," although in this instance there was no control of the possible effects of selective breeding in successive generations (71). One commonly recognized early-experience factor affecting wildness is association with human beings. Lion cubs taken from the mother and reared by hand are tamer than those nursed by the lioness (22).

It has been suggested that severe trauma in infancy can affect emotional reactivity in later life. Young mice from a strain that is highly susceptible to audiogenic seizures when stimulated with the sound of an electric bell were exposed to this stimulus from the fourth day to the seventh day of age. No seizures resulted in these infant animals. However, when they were placed in the same situation at a later age, but not given an auditory stimulus, the "traumatized" mice displayed a higher frequency of eliminative responses than did litter-mate controls (44). These findings were originally interpreted as indicating a general increase in "emotional instability," but the test for emotionality was made in the same environmental setting where traumatization had previously been attempted. It is therefore impossible to rule out the alternative explanation that the results were due to direct and specific conditioning. In a similar investigation rats were used as subjects. Infant animals were ex-

posed to a variety of supposedly traumatic stimuli and tested for emotionality in adulthood. No difference was found between experimental and control groups in an open field test or in susceptibility to audiogenic seizures (42).

In a study of emotionality in dogs, three Scottish terriers were raised as pets with a great deal of human handling, and another three were reared in restricted laboratory cages with a minimum of human contact. As adults, the cage-reared animals differed from the others in that they displayed "freezing" behavior in an unfamiliar environment, hugging the floor and staring forward, as well as a behavioral "anesthesia" to a hypodermic needle (18).

Aggressiveness. As noted earlier, chicks reared in isolation tend to peck other chicks more than do group-reared birds, and the fighting behavior of adult mammals of some species is markedly influenced by certain types of infant experience. Hungry infant mice were trained to fight over food shortly after weaning. As adults, these animals fought over food even when not hungry. Littermates, raised without competitive experience, did not show this behavior except under conditions of starvation (36).

Early defeats in a fighting situation have been shown to influence aggressiveness in later life. Mice of different ages were exposed to and attacked by a second mouse that had been previously trained to fight. The results showed that the younger the mouse was at the time of its initial defeat, the greater was the perseveration of nonaggressive, withdrawal responses in a fighting situation in adulthood (67). In experiments of this type, it is difficult to conclude whether the results are due to the age variable per se, or to the possibility

that a young animal would experience a greater amount of trauma than an older animal.

LEARNING

Many psychologists appear to conceive of learning ability in animals as genetically determined and relatively unmodifiable, but recent findings indicate the untenability of this thesis. Harlow (45) and his co-workers have demonstrated the development of "learning sets" in primates, wherein an animal "learns how to learn," and other studies have shown that stimulus-response learning in rats is significantly facilitated when they have previously learned quite different responses to the training stimuli (64, 74).

Results of a second group of experiments indicate that an animal's performance on special "intelligence tests" is directly related to certain kinds of experience in infancy. In one way or another, most of these experiments are outgrowths of a theory proposed by Hebb (52). In brief, this theory predicts that animals that have had a large amount of perceptual experience early in life will prove better learners than others deprived of such experience. Further, it is predicted that the magnitude of this facilitative effect is, within limits, inversely related to the age at which the perceptual experience is gained.

Evidence for this view has been obtained largely by varying the environment in which the animal matured. Rats reared from 21 to 51 days of age in a small room and not in cages were found to be superior in maze learning to other animals that had been reared in small cages or in "squeeze boxes" (11). In a more carefully controlled study, one group of rats was reared in a large cage, and members of a second group were reared individually in mesh cages which were periodically moved from one position to another in this large cage. A third group was kept in enclosed activity wheels, and a fourth group in enclosed stove-pipe cages. Rats in the first two groups performed significantly better on the Hebb-Williams maze than did the latter two groups. It was concluded that the visual perception in a spacious environment during early life facilitates adult learning ability (62). A further study confirmed the preceding results and added the fact that having early access to pieces of apparatus somewhat similar to parts of the Hebb-Williams maze further increased learning scores on that particular test (35).

Rats reared as pets in the home also show definite superiority in certain learning situations when compared with animals reared in laboratory cages (51). In this connection it should be noted that the mere handling of rats from infancy improves performance on a T maze in adulthood (8), and it is possible that the difference between the home-reared and laboratory-reared animals might be due to this factor.

The age at which the extended environment is experienced is an important factor. Rats reared for a period of time early in life in a large space are clearly superior on the Hebb-Williams maze to animals that are first kept in small cages and then put into the more favorable environment at a later age (62). Supporting the same conclusion in another way is the fact that rats blinded early in life are inferior in maze performance to rats blinded at a later age (51, 62).

One possible explanation for these differences may be that rats reared in a more spacious and heterogeneous environment develop a superior ability to use distance orientation cues in learning. This hypothesis is in accord

with the finding that being raised in a large cage does not improve the rat's ability to learn a simple alley maze (62), or a visual discrimination problem (11). Presumably distance cues are less important in such tests than in more complex measures of the Hebb-Williams type. Positive evidence for the hypothesis is seen in the fact that rotation of the Hebb-Williams maze interferes much more seriously with the performance of rats reared in large than those reared in small cages (35, 62).

Exploratory tendencies may also be involved in the differences under consideration. One of the first studies of this problem showed that the tendency of rats to explore a novel environment in adulthood is increased if the animals have been reared in large instead of small cages (96).

A few experiments have been conducted to extend some of these findings to other species. For example, dogs reared as household pets are described as superior in learning ability to others raised in laboratory cages (18). However, it is impossible to estimate the degree to which the marked emotional differences between the two groups of dogs, as discussed earlier in this review, may have contributed to differences in test performance.

A quite different aspect of infant experience that affects an animal's learning ability in adulthood is the amount of food available. Rats deprived of food in infancy, and thus morphologically stunted, show higher activity scores as adults, and take less time to learn a maze for a food reward than do normally reared animals (3, 4). This finding holds true for mice as well (72). A later study showed that these results may have been due to the fact that the physiological condition of inanition continued to the time of learning in the

experimental animals, rather than to other effects of early experience (10). A perceptual factor may also be involved. Rats which have been deprived of food and water in infancy surpass normal animals in learning the locations of food and water in a maze even though they are satiated at the time of testing (17).

The studies of Harlow (45) on primate learning have shown how the learning performance of monkeys and chimpanzees may be increased by prolonged experience with various aspects of the problems presented. According to Hebb's theory, if this type of experience were gained in infancy it would have an even more marked effect on adult learning capacity, but very few such experiments have been conducted with primates. In one study a home-reared chimpanzee was compared on four simple problems with a laboratory-reared chimpanzee, and with four human children the same age as the animal. The home-reared ape and the children did equally well but the laboratory chimpanzee could only solve one out of the four problems. However, since the solution required imitation of the human experimenter, a requisite which the laboratory-reared animal had had no opportunity to acquire, the interpretation of these results is ambiguous (48).

DISCUSSION

Reviewing the studies in the foregoing pages has left the authors with two basic convictions. The first is that a fuller understanding of the effects of early experience upon subsequent behavior is fundamentally important for a science of psychology and for the broader field of animal biology. A second and equally strong impression is that much if not most of the presently available evidence bearing upon this problem is equivocal

and of undetermined reliability.

Many of the studies cited in this review are primarily of a naturalistic character and some of them verge on the anecdotal. Approximately 10 per cent of the reports contain conclusions based upon the behavior of a single individual. Many of the recent laboratory investigations reveal weaknesses in experimental design, and some of them are marked by lack of sufficient skepticism in theoretical interpretation. We have noted at various points the way in which variables have been confounded without recognition of the fact. Thus, the effects of special conditions of rearing upon intellectual performance have been described in some cases without acknowledgement of the additional effects upon emotional stability which might well influence performance in learning tests. The results of preventing a particular response have been described without considering the confounding effects of deprivation in sensory experience. Natural selection in studies involving successive generations, speed of eating in measures of hoarding, and other variables have not always been subjected to control in the experimental design.

It must be admitted that there are instances in this area where independent control of all the relevant variables is difficult, if not actually impossible. This is especially true with respect to the age variable itself in studies designed to measure the later effects of experimental treatments at different age levels in infancy. Anatomical differences, nutritional requirements, sensory sensitivity, motor development, and previous experience are closely interwoven variables with age, and cannot usually be controlled independently of each other. Any effect ascribed to the different ages at which a treatment is given may be due, not to chronological primacy by itself, but to differences in any one of the above variables or combination of them. This dilemma does not detract from the value of studying such problems. Rather, it means two things: first, that there are stages in the history of every science when investigators are, for a time, forced by the limitations of their knowledge to study the interaction of variables, rather than isolated ones; and second, that such interactions should be seen and labeled as such whenever theoretical conclusions are made.

Because the general thesis that early experience does affect later behavior is so often taken for granted and assumed without further elucidation, it may be clarifying to conclude with some consideration of the basic ideas underlying this general type of research. In what ways does early experience influence later behavior? Broadly speaking, the evidence we have surveyed suggests three answers to this question.

Persistence in adult behavior of habits formed in early life. The persistence of acquired behavior through time is, of course, a general phenomenon throughout the life of an animal, but there are several reasons why habits acquired early in life might be especially persistent. When a very young animal learns a response to a stimulus situation, the occurrence of that response each time the situation presents itself may prevent the acquisition of other types of behavior which could compete with the original one. This appears to have been the case in some of the studies on pecking in chicks, in which the acquisition of an unnatural eating response to a spoonful of food prevented the development of the normal pecking reaction (96). The same explanation seems applicable to the finding that male rats which, in infancy, have formed habits of respond-

ing playfully to receptive females, never display completely normal sexual reactions after they become mature (66).

An additional reason why habits learned early in life are sometimes retained better than those which are acquired later involves differences in motivation with age. It is possible that certain physiological needs and their corresponding "drives" are more intense in infancy than in adulthood. Younger rats, for example, will make many more grid crossings to food than older rats tested after the same duration of food deprivation (81). Rats starved just after weaning are therefore probably hungrier and have a more intense experience than rats starved at a later period. As a result, the effects of early experience might be expected to persist because of the intensity of the motivation under which the experience originally occurred, while the effects of later experience would not persist. It is conceivable that Hunt's first results showing the greater effect of early than late deprivation of food on adult hoarding of food (59) can be interpreted in this way. A similar case could be made for the more lasting effects of early trauma. Because of the young animal's unfamiliarity with its environment, and its lack of a learned repertoire of fear-reducing responses, fear in infancy is probably of a greater intensity than similarly produced fear later in life. It is known, for example, that incidence of seizure reactions to air blasts is a decreasing function of age in rats (30, 34). In the same manner, defeat in a fight, as occurred in Kahn's study of mice (67), may be a more intense experience for a young animal than for an older one; and for this reason, its traumatic effects might last longer and have a greater influence on subsequent behavior.

Early perceptual learning affecting adult behavior. Quite distinct from the theory of retention in adulthood of the same overt responses learned in infancy is the hypothesis that certain types of early experience influence later behavior by structuring the individual's perceptual capacities. Practically all the studies we have cited in the fields of sensation and learning are based upon this view. Mention has already been made of Hebb's analysis of the development of behavior. Early learning is conceived to consist largely of the establishment of "perceptual elements" ("phase sequences," neurologically) which serve as the basis of learning later in life (52). It is far too early to evaluate the explanatory value of this theory, but several studies have shown that absence of normal sensory or motor experience in infancy can severely restrict the reactive abilities of adult animals.

Critical periods in development. A third idea which underlies many researches involves the concept of specific stages in ontogeny during which certain types of behavior normally are shaped and molded for life. This hypothesis may have been suggested by certain findings in experimental embryology. If at a certain stage in the development of the frog embryo, part of the presumptive belly wall is transplanted to the region where the nervous system is later to develop, the transplanted tissue develops into a portion of the neural tube—the future of the cells thus being determined in part by their environment, a process which Spemann (121) called induction. At a later stage, however, the cells lose their susceptibility to such environmental influence. Similarly, there may be relatively brief stages in the behavioral development

of animals during which the future of certain aspects of behavior is strongly affected by contemporary environmental influences that have no such effect at other points along the life line.

Applying this theoretical viewpoint to the behavioral development of birds, Lorenz (78) concluded that there are specific and restricted periods during which the stimuli which will thereafter evoke certain instinctive responses are permanently determined. After the critical period has passed, the environment cannot alter the nature of the effective stimulus. Studies on imprinting suggest that this may be true for the filial responses of ground-nesting birds, in which case the critical period is restricted to the first 24 hours after hatching (1, 29).

More recently, the idea of critical periods has been applied to dogs. Scott and Marston (114) divide the development of the puppy into sequential periods according to the kinds of behavior that are possible. First is the neonatal period, lasting approximately ten days, when the animal is both blind and deaf, and therefore somewhat isolated from the environment. A transition period follows to three weeks of age, at which point conditioned responses are said to become possible fairly suddenly (39). Then comes a socialization phase, lasting to ten weeks of age. This is described as a critical period affecting the future social adjustment of the individual (113). It is followed by the juvenile period, and, after sexual maturity, the adult stage.

There are several obvious similarities between the hypothesized critical periods in animals and various theories for the development of human behavior. For example, Freud's concept of discrete stages in psychosexual development and the lasting effects of fixation at one level or another calls to mind Levy's demonstration that interference with normal oral responses in a period when they predominate in an animal's behavior may influence later behavior of the same kind. And Murphy's use of the concept of canalization (90), which is regarded as an irreversible learning process whereby some needs come to be satisfied only or predominately by certain, specific stimuli, bears a resemblance to Lorenz's theory of imprinting. Descriptions of the developmental stages in puppies are somewhat reminiscent of Piaget's account of the progressive development of social behavior in young children (99). It is tempting to suggest the existence of similarities in underlying mechanisms, but this would be unwarranted on the basis of present knowledge.

In listing three possible answers to the question of how early experience can affect later behavior, we mean to imply neither that they are mutually independent, nor that they are together exhaustive. Rather they appear as currents of thought along which research in this area has progressed.

In conclusion, it can be said that the longitudinal approach to the study of animal life is obviously more beset with methodological problems than studies concerned with only one portion of the life span. Experimental control in this area is more difficult to realize, because the animals, by the very nature of the problem, are exposed to numerous variables, some of which may not yet be recognized or capable of isolation. These difficulties, however, are more than offset by the general interest and wide significance of the problems that may be studied.

REFERENCES

1. ALLEY, R., & BOYD, H. Parent-young recognition in the coot. *Ibis*, 1950, 92, 46–51.

2. ALTMANN, M. Social behavior of elk, *Cervus Canadensis* Nelsoni, in the Jackson Hole area of Wyoming. *Behaviour*, 1952, 4, 116–143.

3. ANDERSON, J. E., & SMITH, A. H. The effect of quantitative and qualitative stunting upon maze learning in the white rat. *J. comp. Psychol.*, 1926, 6, 337–359.

4. ANDERSON, J. E., & SMITH, A. H. The relation of performance to age and nutritive condition in the white rat. *J. comp. Psychol.*, 1932, 13, 409–446.

5. ARISTOTLE. *History of animals.* (Trans. by Richard Cresswell.) London: George Bell, 1891.

6. BEACH, F. A. The neural basis of innate behavior: I. Effects of cortical lesions upon the maternal behavior pattern in the rat. *J. comp. Psychol.*, 1937, 24, 393–436.

7. BEACH, F. A. Comparison of copulatory behavior of male rats raised in isolation, cohabitation, and segregation. *J. genet. Psychol.*, 1942, 60, 137–142.

8. BERNSTEIN, L. A note on Christie's: "Experimental naïveté and experiential naïveté." *Psychol. Bull.*, 1952, 49, 38–40.

9. BETHE, A. Dürfen wir den Ameisen und Bienen psychische Qualitäten zuschreiben? *Pflüg. Arch. ges. Physiol.*, 1898, 70, 15–100.

10. BIEL, W. C. The effect of early inanition upon maze learning in the albino rat. *Comp. Psychol. Monogr.*, 1938, 15, No. 2.

11. BINGHAM, W. E., & GRIFFITHS, W. J., JR. The effect of different environments during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, 45, 307–312.

12. BIRD, C. Maturation and practice; their effects upon the feeding reactions of chicks. *J. comp. Psychol.*, 1933, 16, 343–366.

13. BRATTGÅRD, S. The importance of adequate stimulation for the chemical composition of retinal ganglion cells during early post-natal development. *Acta Radiologica*, 1952, Suppl. 96.

14. BREDER, C. M., & RASQUIN, P. Comparative studies in the light sensitivity of blind characins from a series of Mexican caves. *Bull. Amer. Mus. Nat. Hist.*, 1947, 89, 325–351.

15. BRÜCKNER, G. H. Untersuchungen zur Tiersoziologie, insbesondere zur Auflösung der Familie. *Zsch. f. Psychol.*, 1933, 128, 1–110.

16. CARMICHAEL, L. The development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.*, 1926, 34, 34–47.

17. CHRISTIE, R. The effect of some early experiences in the latent learning of adult rats. *J. exp. Psychol.*, 1952, 43, 281–288.

18. CLARKE, R. S., HERON, W., FETHERSTONHAUGH, M. L., FORGAYS, D. G., & HEBB, D. O. Individual differences in dogs: preliminary report on the effects of early experience. *Canad. J. Psychol.*, 1951, 5, 150–156.

19. COLLIAS, N. E. Social life and the individual among vertebrate animals. *Ann. N. Y. Acad. Sci.*, 1950, 51, 1074–1092.

20. COLLIAS, N. E. The development of social behavior in birds. *Auk*, 1952, 69, 127–159.

21. CONRADI, E. Song and call notes of English sparrows when reared by canaries. *Amer. J. Psychol.*, 1905, 16, 190–198.

22. COOPER, J. B. An exploratory study on African lions. *Comp. Psychol. Monogr.*, 1942, 17, No. 7.

23. CRAIG, W. The voices of pigeons regarded as a means of social control. *Amer. J. Sociol.*, 1908, 14, 86–100.

24. CRAIG, W. Male doves reared in isolation. *J. Anim. Behav.*, 1914, 4, 121–133.

25. CRUZE, W. W. Maturation and learning in chicks. *J. comp. Psychol.*, 1935, 19, 371–409.

26. CUSHING, J. E., JR. An experiment on olfactory conditioning in *Drosophila guttifera*. *Proc. Nat. Acad. Sci.*, 1941, 27, 496–499.

27. DARLING, F. F. *A herd of red deer.* London: Oxford Univer. Press, 1937.

28. ELLIS, PEGGY E. Social aggregation and gregarious behavior in hoppers of *Locusta migratoria migratorioides* (R. & F.). *Behaviour*, 1953, 5, 225–258.

29. FABRICIUS, E. Zur Ethologie junger Anatiden. *Acta Zool. Fenn.*, 1951, 68, 1–175.

30. FARRIS, E. F., & YEAKEL, E. H. Sex and increasing age as factors in the frequency of audiogenic seizures in albino rats. *J. comp. Psychol.*, 1942, 34, 75–78.

31. FAURE, J. C. The phases of locusts in

South Africa. *Bull. ent. Res.*, 1932, **23**, 293–405.

32. FIELDE, ADELE M. Artificial mixed nests of ants. *Biol. Bull.*, 1903, 5, 320–325.

33. FIELDE, ADELE M. Power of recognition among ants. *Biol. Bull.*, 1904, **7**, 227–250.

34. FINGER, F. W. Factors influencing audiogenic seizures in the rat. II. Heredity and age. *J. comp. Psychol.*, 1943, **35**, 227–232.

35. FORGAYS, D. G., & FORGAYS, JANET W. The nature of the effect of free-environmental experience in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 322–328.

36. FREDERICSON, E. Competition: the effects of infantile experience upon adult behavior. *J. abnorm. soc. Psychol.*, 1951, 46, 406–409.

37. FREDERICSON, E. The wall-seeking tendency in three inbred mouse strains (*Mus musculus*). *J. genet. Psychol.*, 1953, **82**, 143–146.

38. FREUD, S. Three essays on the theory of sexuality. London: Imago Publishing Co., Ltd., 1949.

39. FULLER, J. L., EASLER, C. A., & BANKS, E. M. Formation of conditioned avoidance responses in young puppies. *Amer. J. Physiol.*, 1950, **160**, 462–466.

40. GOODMAN, L. The effect of total absence of function upon the optic system of rabbits. *Amer. J. Physiol.*, 1932, 100, 46–63.

41. GOODWIN, D. Some abnormal sexual fixations in birds. *Ibis.*, 1948, 90, 45–48.

42. GRIFFITHS, W. J., & STRINGER, W. F. The effects of intense stimulation experienced during infancy on adult behavior in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 301–306.

43. HAGEN, W. Ist des Gesang des Buchfinken angeboren oder erlernt? *Ornith. Monatsber.*, 1923, **31**, 52–54.

44. HALL, C. S., & WHITEMAN, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.*, 1951, **44**, 61–66.

45. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

46. HARRISON, J. W. H. Experiments on the egg-laying habits of the sawfly *Pontania salicis* Chr. *Proc. roy. Soc.*, 1927, **101B**, 115.

47. HARRISON, R. G. An experimental study of the relation of the nervous system to the developing musculature in the embryo of the frog. *Amer. J. Anat.*, 1904, **3**, 197–220.

48. HAYES, K. J., & HAYES, CATHERINE. The intellectual development of a home-raised chimpanzee. *Proc. Amer. Phil. Soc.*, 1951, **95**, 105–109.

49. HEBB, D. O. The innate organization of visual activity: I. Perception of figures by rats reared in darkness. *J. genet. Psychol.*, 1937, **51**, 101–126.

50. HEBB, D. O. The innate organization of visual activity: II. Transfer of response in the discrimination of brightness and size by rats reared in total darkness. *J. comp. Psychol.*, 1937, **24**, 277–299.

51. HEBB, D. O. The effects of early experience on problem-solving at maturity. *Amer. Psychologist*, 1947, **2**, 306–307. (Abstract)

52. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

53. HEDIGER, H. *Wild animals in captivity.* London: Butterworth's, 1950.

54. HEINROTH, O., & HEINROTH, M. *Die Vögel Mitteleuropas.* Berlin: Lichterfelde, 1924.

55. HESS, E. H. Development of the chick's responses to light and shade cues of depth. *J. comp. physiol. Psychol.*, 1950, **43**, 112–122.

56. HICKEY, J. Eastern population of the duck hawk. *Auk*, 1942, 59, 176–204.

57. HILGARD, E. Experimental approaches to psychoanalysis. In E. Pumpian-Mindlin (Ed.), *Psychoanalysis as science.* Stanford: Stanford Univer. Press, 1952. Pp. 3–45.

58. HUDSON, W. H. *The naturalist in La Plata.* London: Chapman and Hall, 1892.

59. HUNT, J. McV. The effects of infant feeding-frustration upon adult hoarding in the albino rat. *J. abnorm. soc. Psychol.*, 1941, **36**, 338–360.

60. HUNT, J. McV., SCHLOSBERG, H., SOLOMON, R. L., & STELLAR, E. Studies of the effects of infantile experience on adult behavior in rats. I. Effects of infantile feeding-frustration on adult hoarding. *J. comp. physiol. Psychol.*, 1947, 40, 291–304.

61. HUXLEY, J. *Evolution: the modern synthesis.* London: Harper, 1942.

62. HYMOVITCH, B. The effects of experimental variations on problem solving in the rat. *J. comp. physiol. Psychol.*, 1952, **45**, 313–320.

63. JACOBSEN, C. F., JACOBSEN, MARION M., & YOSHIOKA, J. G. Development of an infant chimpanzee during her first

year. *Comp. Psychol. Monogr.*, 1932, 9.

64. JAYNES, J. Learning a second response to a cue as a function of the magnitude of the first. *J. comp. physiol. Psychol.*, 1950, **43**, 398–408.

65. JENKINS, M. The effect of segregation on the sex behavior of the white rat as measured by the obstruction method. *Genet. Psychol. Monogr.*, 1928, **3**, 457–471.

66. KAGAN, J., & BEACH, F. A. Effects of early experience on mating behavior in male rats. *J. comp. physiol. Psychol.*, 1953, **46**, 204–208.

67. KAHN, M. W. The effect of severe defeat at various age levels on the aggressive behavior of mice. *J. genet. Psychol.*, 1951, **79**, 117–130.

68. KATZ, D. *Animals and men*. New York: Longmans, Green, 1937.

69. KELLOGG, W. N., & KELLOGG, LUELLA A. *The ape and the child*. New York: Mc-Graw-Hill, 1933.

70. KENNEDY, J. S. The behaviour of the Desert Locust in an outbreak centre. *Trans. roy ent. Soc. Lond.*, 1939, **89**, 385–542.

71. KING, H. D. Life processes in gray Norway rats during 14 years in captivity. *Amer. Anat. Mem.*, 1939, 17, 1–72.

72. KOCH, A. M., & WARDEN, C. J. The influence of quantitative stunting on learning ability in mice. *J. genet. Psychol.*, 1936, **48**, 215–217.

73. KÖHLER, W. *The mentality of apes*. New York: Harcourt, Brace, 1925.

74. LAWRENCE, D. H. The acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

75. LASHLEY, K. S., & RUSSELL, J. T. The mechanism of vision. XI. A preliminary test of innate organization. *J. genet. Psychol.*, 1934, **45**, 136–144.

76. LEVY, D. M. Experiments on the sucking reflex and social behavior in dogs. *Amer. J. Orthopsychiat.*, 1934, 4, 203–224.

77. LEVY, D. M. On instinct-satiation: an experiment on the pecking behavior of chickens. *J. genet. Psychol.*, 1938, **18**, 327–348.

78. LORENZ, K. Der Kumpan in der Umwelt des Vogels. Der Artgenosse als auslösendes Moment sozialer Verhaltungsweisen. *J. Orn. Lpz.*, 1935, **83**, 137–213; 289–413.

79. McKELVEY, R. K., & MARX, M. H. Effects of infantile food and water deprivation on adult hoarding in the rat.

J. comp. physiol. Psychol., 1951, 44, 423–430.

80. MAIER, N. R. F., & SCHNEIRLA, T. V. *Principles of animal psychology*. New York: McGraw-Hill, 1935.

81. MARGOLIN, S. E., & BUNCH, M. E. The relationship between age and the strength of hunger motivation. *Comp. Psychol. Monogr.*, 1940, No. 16.

82. MARX, M. H. A stimulus-response analysis of the hoarding habit in the rat. *Psychol. Rev.*, 1950, 57, 80–93.

83. MARX, M. H. Infantile deprivation and adult behavior in the rat: retention of increased rate of eating. *J. comp. physiol. Psychol.*, 1952, **45**, 43–49.

84. MAYR, E. *Systematics and the origin of species*. New York: Columbia Univer. Press, 1942.

85. METFESSEL, M. Roller canary song produced without learning from external sources. *Science*, 1935, **81**, 470.

86. MORGAN, C. L. *Animal behavior*. London: Arnold, 1900.

87. MOSELEY, D. The accuracy of the pecking response in chicks. *J. comp. Psychol.*, 1925, **5**, 75–97.

88. MOWRER, O. H. Maturation vs. "learning" in the development of vestibular and optokinetic nystagmus. *J. genet. Psychol.*, 1936, **48**, 383–404.

89. MURIE, A. The wolves of Mt. McKinley. *Nat. Park Ser., U. S. Dep. of Int.*, 1944, Fauna Series, No. 5.

90. MURPHY, G. *Personality*. New York: Harper, 1947.

91. NISSEN, H. W. Social behavior in primates. In C. P. Stone (Ed.), *Comparative psychology*. (3rd Ed.) New York: Prentice-Hall, 1951. Pp. 423–457.

92. NISSEN, H. W., CHOW, K. L., & SEMMES, JOSEPHINE. Effects of restricted opportunity for tactual, kinesthetic, and manipulative experience on the behavior of a chimpanzee. *Amer. J. Psychol.*, 1951, 64, 485–507.

93. NOBLE, G. K., & CURTIS, B. Sexual selection in fishes. *Anat. Rec.*, 1935–1936, 64, 84–85.

94. NOBLE, G. K., & CURTIS, B. The social behavior of the jewel-fish, *Hemichromis bimaculatus* Gill. *Bull. Amer. Mus. Nat. Hist.*, 1939, 76, 1–46.

95. PADILLA, S. G. Further studies on delayed pecking in chicks. *J. comp. Psychol.*, 1935, 20, 413–443.

96. PATRICK, J. R., & LAUGHLIN, R. M. Is the wall-seeking tendency in the white rat an instinct? *J. genet. Psychol.*, 1934, **44**, 378–389.

97. Pattie, F. A., Jr. The gregarious behavior of normal chicks and chicks hatched in isolation. *J. comp. Psychol.*, 1936, 21, 161–178.

98. Phillips, J. C. The Hawaiian linnet, *Carpodacus mutans* Grinnell. *Auk*, 1912, 29, 336–338.

99. Piaget, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.

100. Pride, A. Notes on the habits of the rhea. *Proc. roy. phys. Soc., Edin.*, 1915, 19.

101. Räber, H. Analyse des Balzverhaltens eines domestizierten Truthahns (Meleagris). *Behaviour*, 1948, 1, 237–266.

102. Ramsey, A. O. Familial recognition in domestic birds. *Auk*, 1951, 68, 1–16.

103. Rasmussen, E. W. Wildness in the rat. *Acta Psychol.*, Hague, 1939, 4, 295–304.

104. Ribble, M. A. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 621–651.

105. Riesen, A. H. The development of visual perception in man and chimpanzee. *Science*, 1949, 106, 107–108.

106. Riesen, A. H. Arrested vision. *Sci. Amer.*, 1950, 183, 16–19.

107. Riesen, A. H. Post-partum development of behavior. *Chicago Med. School Quart.*, 1951, 13, 17–24.

108. Riess, B. F. The isolation of factors of learning and native behavior in field and laboratory studies. *Ann. N. Y. Acad. Sci.*, 1950, 51, 1093–1102.

109. Ross, S. Sucking frustration in neonate puppies. *J. abnorm. soc. Psychol.*, 1951, 46, 142–149.

110. Sadovnikova-Koltzova, M. P. Genetic analysis of temperament of rats. *J. exp. Zool.*, 1926, 45, 301–318.

111. Salt, G. The effects of hosts upon their insect parasites. *Biol. Rev.*, 1941, 14, 239–264.

112. Scott, J. P. Social behavior, organization, and leadership in a small flock of domestic sheep. *Comp. Psychol. Monogr.*, 1945, 18, No. 4.

113. Scott, J. P., Fredericson, E., & Fuller, J. L. Experimental exploration of the critical period hypothesis. *Personality*, 1951, 1, 162–183.

114. Scott, J. P., & Marston, M. V. Critical periods affecting the development of normal and maladjustive social behavior in puppies. *J. genet. Psychol.*, 1950, 77, 25–60.

115. Scott, W. E. D. Data on songbirds. *Science*, 1902, 15, 178–181.

116. Scott, W. E. D. The inheritance of song. *Science*, 1904, 19, 957–959.

117. Shepherd, J. F., & Breed, F. S. Maturation and use in the development of an instinct. *J. Anim. Behav.*, 1913, 3, 274–285.

118. Siegel, A. I. Deprivation of visual form definition in the ring dove. I. Discriminatory learning. *J. comp. physiol. Psychol.*, 1953, 46, 115–119.

119. Small, W. S. Notes on the psychic development of the young white rat. *Amer. J. Psychol.*, 1899, 11, 80–100.

120. Speidel, C. Milwaukee's polar bears. *Parks and Recreation*, 1949, 32, 235–237.

121. Spemann, H. *Embryonic development and induction.* New Haven: Yale Univer. Press, 1938.

122. Stone, C. P. The initial copulatory response of female rats reared in isolation from the age of 20 days to age of puberty. *J. comp. Psychol.*, 1926, 6, 78–83.

123. Thorpe, W. H. Further experiments on olfactory conditioning in a parasitic insect. The nature of the conditioning process. *Proc. roy. Soc.*, 1938, 126B, 370–397.

124. Thorpe, W. H. Further studies on preimaginal olfactory conditioning in insects. *Proc. roy. Soc.*, 1939, 127B, 424–433.

125. Thorpe, W. H. Types of learning in insects and other arthropods. *Brit. J. Psychol.*, 1943, 33, 220–234; 1944, 34, 20–31; 66–76.

126. Thorpe, W. H., Some problems of animal learning. *Proc. Linn. Soc. Lond.*, 1944, 156, 70–83.

127. Thorpe, W. H. The learning abilities of birds. Part 2. *Ibis.*, 1951, 93, 252–296.

128. Thorpe, W. H., & Jones, F. G. W. Olfactory conditioning in a parasitic insect and its relation to the problems of host selection. *Proc. roy. Soc.*, 1937, 124B, 56–79.

129. Uvarov, B. P. *Locusts and grasshoppers.* London: Imperial Bureau of Entomology, 1928.

130. Volkonsky, M. Sur la photo-akinèse des acridiens. *Arch. Inst. Pasteur Algér.*, 1939, 17, 194–220.

131. Wolfe, J. B. An exploratory study of food-storing in rats. *J. comp. Psychol.*, 1939, 28, 97–108.

132. Yerkes, R. M. *The dancing mouse.* New York: Macmillan, 1907.

SEQUENTIAL ANALYSIS IN PSYCHOLOGICAL RESEARCH

DONALD W. FISKE AND LYLE V. JONES[1]
The University of Chicago

The purpose of this paper is to introduce sequential analysis to psychological researchers, and more incidentally, to discuss the operating characteristic function for statistical tests in general. Sequential analysis is not presented adequately in the statistical texts most commonly consulted by psychologists and has rarely been used in psychological research. Yet it is a statistical method that has the important advantage that it minimizes the average number of observations required to reach a specific statistical decision; probably it will be most useful to research on personality and other topics where the cost per unit (i.e., per observation or pair of observations) is high. The approach to statistical problems in terms of the operating characteristic or power of a test also has received little attention in textbooks on psychological statistics (cf. **8**, pp. 65–67). This approach, which is important in sequential analysis, has general relevance to statistical tests of hypotheses.

Sequential analysis is fully presented in a book written by its principal inventor, Wald (**14**). A brief technical presentation can be found in Mood (**11**). A more definitive and detailed exposition, presented in the context of sampling inspection, is also available (**3**). The chapter discussing it in Dixon and Massey (**5**, ch. 18), while presenting sequential procedures in somewhat different form, may be the most readily available introduction for psychological investigators.

The few relevant papers using sequential analysis have essentially been demonstrations of the method as applied to test construction and test administration. Cowden (**4**) gave a final course examination sequentially, with each student being required to answer successive sets of questions until a decision could be reached concerning his achievement. Moonan (**12**) suggests possible uses of the method for problems involving the optimal number of items or for spotting extreme students, but concludes that the method is usually not worth while in education, where the expense of collecting data is relatively independent of the number of items used. The method was applied by Kimball (**10**) to the problem of checking and rescoring test papers. The application of sequential sampling to item selection is shown by Walker (**15**), who, with Cohen (**16**), provides facilitating tables. Schmid (**13**) offers another approach to the same problem. Good agreement is reported by Anastasi (**1**) between decisions about item selection based on sequential analysis and those based on correlations for a total sample.

THE OPERATING CHARACTERISTIC FUNCTION

The typical elementary text in psychological statistics defines the probability of errors of the first kind, α, and the probability of errors of the second kind, β. Errors of the first kind (Type I errors) are errors of rejecting, on the basis of sample observations, a hypothesis, H_0, when

[1] The authors are indebted to Professor W. Allen Wallis and Dr. Esther Seiden for their valuable criticisms of an early draft of this paper.

H_0 is, in the population, true; errors of the second kind (Type II errors) are errors of accepting H_0, when in the population an alternative hypothesis, H_1, is true. Having presented such definitions, it is the rare text in psychological statistics which further discusses the importance of errors of the second kind. Rather, discussions of actual statistical tests of hypotheses emphasize confidence levels, α errors, to the exclusion of the probability of errors of the second kind, β errors.

The operating characteristic (OC) curve, for any statistical test of hypotheses, shows the relationship between the probability of accepting a particular hypothesis and the true value of the parameter under consideration. The height of the OC curve at a given parametric value, μ, represents the probability of accepting the specified hypothesis when the observed sample of observations has been randomly selected from a population with the parameter μ. The use of the OC curve is exemplified in Fig. 1.

Let us suppose that we are drawing a large random sample from a normally distributed population. Let us choose a level of significance, $\alpha = .05$, for our test. In Fig. 1 are exemplified two OC curves, one for each of the following two tests:

1. We wish to test the hypothesis that the population mean is zero ($H_0: \mu = 0$) against the alternative hypothesis that μ differs from zero. The height of curve A, in Fig. 1, gives us the probability of accepting the hypothesis H_0 for all possible parametric values of μ. For example, if the true population mean is zero, the probability that we will accept this hypothesis is .95, or, in general, $1 - \alpha$. If the true population mean differs from zero by three standard
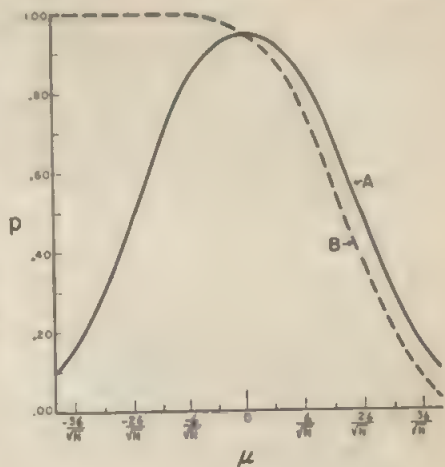
FIG. 1. ILLUSTRATIVE OC CURVES FOR ONE-TAILED AND TWO-TAILED TESTS

errors of the mean, the probability that we would accept H_0 is reduced to approximately .16. This value is equal to β, the error of the second kind, for the particular alternative hypothesis $\mu = 3\sigma/\sqrt{N}$. In general, the height of the OC curve is β for any point along the abscissa other than that point included in the specified hypothesis, H_0.

2. We wish to test the hypothesis $\mu = 0$ against the alternative hypothesis $\mu > 0$. The probability of accepting this hypothesis, as a function of the true parametric mean, is given by the height of OC curve B in Fig. 1. Again that height at $\mu = 0$ is .95 or $1 - \alpha$. At $\mu = 3\sigma/\sqrt{N}$ that height is .08. For the particular alternative hypothesis $\mu = 3\sigma/\sqrt{N}$, the β error for this test is half the magnitude of the β error for test 1 above. For this alternative hypothesis, this test is more "powerful" than test 1. In general, the "power" of a statistical test is defined as $1 - \beta$ for any particular alternative hypothesis; while from an OC curve, one reads the magnitude of β; from a "power curve," one reads the magnitude of $1 - \beta$.

If, in this hypothetical problem, we were primarily interested in guarding against the alternative hypothesis $\mu = 3\sigma/\sqrt{N}$, then we would choose test 2 rather than test 1, thus reducing the probability of an error of the second kind. Such a choice would preserve the confidence level, $\alpha = .05$, but would maximize the power of our test. Often, in designing statistical tests of hypotheses, we have a choice among a number of alternative tests. Our choice can be optimally resolved if we know the OC function for each of the tests, for then we can choose that test which maximizes the power (minimizes β) for a particular alternative in which we may be especially interested.

AN EXAMPLE OF THE SEQUENTIAL METHOD[2]

The general approach used in sequential analysis may now be illustrated with a hypothetical example. Suppose that we wished to evaluate a new psychotherapeutic technique. Let us assume that an acceptable criterion for therapeutic success is available and that it has been established that the best one of the conventional techniques is successful in 60 per cent of cases treated. We are interested in determining whether or not the new technique is more successful than the conventional one.

To make a test on these bases, we must be able to observe results of the new technique which may be considered random representations of a population of results from this technique, and which can be reliably scored as successes or nonsuccesses. Let us assume that these requirements have been met. Our sequential plan will consist of examining in sequence single results or groups of

results, then of making one of three possible decisions: (a) that the new technique is no better than the conventional one, (b) that the new technique is better than the conventional one, or (c) that we do not have sufficient evidence to make either decision without too great a risk of error. If the third decision is taken, another result or group of results is observed, and on the basis of this additional observation, the three decisions are reconsidered. Sampling proceeds in this manner until we have a basis for making either decision a or decision b.

In general, the plan for sequential sampling can be expressed in pictorial form, as in Fig. 2. For this binomial problem, the cumulative proportion of successes is plotted against the total number of observations. So long as the plotted line remains in the region bounded by the two parallel straight lines, sampling continues. Should the plotted line reach the upper straight line, the hypothesis under test is rejected in favor of the
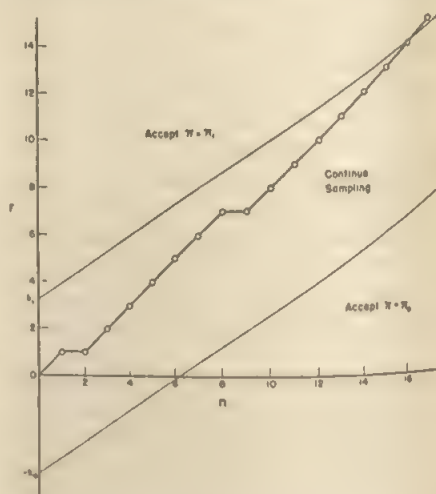

FIG. 2. A SEQUENTIAL SAMPLING PLAN FOR TESTING HYPOTHESIS ABOUT PROPORTION AGAINST ONE-SIDED ALTERNATIVE

---

[2] The succeeding presentation closely parallels that in (3).

alternative. Should the plotted line reach the lower straight line, the hypothesis under test is accepted. The slope and the intercepts of the two parallel lines are completely predetermined by characteristics of the statistical test.

Now let us return to our example. From the formulation of the problem, one of the hypotheses in which we are interested is obvious, i.e., that the new technique produces no more successes than the old one. In formal terms, we wish to test the hypothesis

$$H_0: \pi = \pi_0 = .60,$$

where π is the proportion of successes in the population from which our sample is drawn. The choice of an alternative hypothesis depends upon what we mean by "better than." In this case, we might decide that if the new technique is better than the old, as evidenced by a 15 per cent increase in successes, such a difference would be of considerable importance; we might also decide that a less dramatic difference than this might be of questionable practical importance. On the basis of such considerations, we would choose a second hypothesis

$$H_1: \pi = \pi_1 = .75.$$

We now choose a risk α of rejecting H_0 when H_0 is true, and a risk β of accepting H_0 when H_1 is true. In our example, we might choose to make β the smaller risk of the two, for it would be a costly mistake to conclude that the new technique was no better than the old, when actually the new technique produced 75 per cent successes as against 60 per cent for the old. We might then choose

$$\alpha = .10,$$

and

$$\beta = .05.$$

Having chosen the four values H_0, H_1, α, and β, we have completely determined the characteristics of our sequential test and are able to consider the OC curve for the test. To draw a rough sketch of a sequential OC curve, we are able to make use of the following five points, where L is the height of the OC curve:

a. $L(\pi = 0) = 1;$

b. $L(\pi = \pi_0) = 1 - \alpha;$

c. $L(\pi = s) = \dfrac{k_1}{k_0 + k_1};$

d. $L(\pi = \pi_1) = \beta;$

e. $L(\pi = 1) = 0;$

where,

$$s = \frac{\log\left(\dfrac{1 - \pi_0}{1 - \pi_1}\right)}{\log\left[\dfrac{\pi_1}{\pi_0}\left(\dfrac{1 - \pi_0}{1 - \pi_1}\right)\right]}, \qquad [1]$$

$$k_0 = \frac{\log\dfrac{1 - \alpha}{\beta}}{\log\left[\dfrac{\pi_1}{\pi_0}\left(\dfrac{1 - \pi_0}{1 - \pi_1}\right)\right]}, \text{ and } [2]$$

$$k_1 = \frac{\log\dfrac{1 - \beta}{\alpha}}{\log\left[\dfrac{\pi_1}{\pi_0}\left(\dfrac{1 - \pi_0}{1 - \pi_1}\right)\right]}. \qquad [3]$$

A number of additional points on the OC curve can be found; for a thorough explanation of methods of determining the OC curve, the reader is referred to the report of the Statistical Research Group, Columbia University (3).

In the present example, the OC curve is as sketched in Fig. 3. We are now able to visualize the characteristics of our sequential test, in terms of the OC curve, and to decide whether
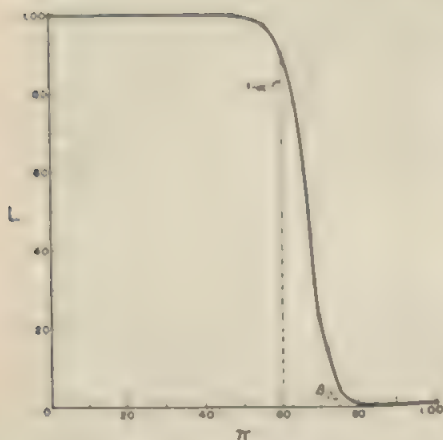
FIG. 3. THE SEQUENTIAL OC CURVE
FOR BINOMIAL EXAMPLE

these characteristics are satisfactory for our purposes. If they are not, we alter H_0 or H_1 or α or β in such a manner as to make the test more suitable. If the characteristics are acceptable, we proceed further in determining the sampling plan.

Referring to Fig. 2, the decision regions are completely defined by a a value for the slope of the two parallel lines, and by an intercept value for each of the two lines. In general, the equations of the two parallel lines may be written

$$r_0 = -k_0 + sn \qquad [4a]$$

$$r_1 = k_1 + sn, \qquad [4b]$$

where r stands for the number of successes, n the cumulative number of observations, s the slope, and $-k_0$ and k_1 for the intercepts. It can be shown that s, k_0, and k_1 are the values defined previously in equations [1], [2], and [3] (3, sec. B.072). For the present example, $k_0 = 4.170$, $k_1 = 3.248$, and $s = .678$. Our sequential plan is that represented by the two parallel lines of Fig. 2.

To illustrate an application of sequential analysis to a binomial problem, consider the following sampling experiment. Observations are selected from a table of random numbers. Assuming a population π of .80, we record any digit from 1 to 8 a success (a "1"); digits 9 and 0 are recorded as failures ("0"). A sequence of such observations has been made, and the results are presented in the order observed:

1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1.

This sequence of observations is graphed in Fig. 2; since the plotted line reaches the upper straight line on the sixteenth observation, sampling stops at that point, and we reject H_0 in favor of H_1. We conclude it to be more likely that $\pi = .75$ than that $\pi = .60$, the confidence in the conclusion being specified by the probability of Type I error, $\alpha = .10$.

THE AVERAGE SAMPLE SIZE

In addition to the OC curve for a sequential test, there is another curve of considerable interest. This is a curve of average sample size (or average sample number, ASN) as a function of the true population parameter. It is useful to refer to the ASN curve to determine the expected cost, in terms of sample size, of achieving that assurance of a correct decision selected from the OC function.

The average sample size depends upon α and β; the smaller these risks, the greater the size that the sample will usually be before a decision is reached. But the ASN also depends upon the hypothesized values π_0 and π_1, and upon the population proportion of successes, π. For given values of α, β, π_0, and π_1 we can write the ASN as a function of π. If π corresponds to the hypothesized proportion, π_0, we have

$$\text{ASN } (\pi = \pi_0) = \frac{(1 - \alpha)k_0 - \alpha k_1}{s - \pi_0}.$$

If π corresponds to π_1,

$$\text{ASN} (\pi = \pi_1) = \frac{(1 - \beta)k_1 - \beta k_0}{\pi_1 - s},$$

where s, k_0, and k_1 are those values given in equations [1], [2], and [3]. These two values, ASN $(\pi = \pi_0)$ and ASN $(\pi = \pi_1)$, are larger than any average sample size necessary to make a decision if π is not between π_0 and π_1.

Three additional points on the ASN curve may be found to be:

$$\text{ASN} (\pi = 0) = \frac{k_0}{s},$$

$$\text{ASN} (\pi = s) = \frac{k_0 k_1}{s(1 - s)}.$$

$$\text{ASN} (\pi = 1) = \frac{k_1}{1 - s}.$$

Thus, just as we were able to sketch the OC curve from five points, given any sequential plan, we are able to sketch the ASN curve from five points. The ASN curve for our earlier example is given in Fig. 4. For any sequential plan it is not difficult to solve for a larger number of points on the ASN curve (3).

FIG. 4. THE ASN CURVE FOR BINOMIAL EXAMPLE

SOME OTHER SEQUENTIAL TESTS

Requirements regarding untested parameters. The simpler forms of sequential test assume that either the variance or the mean is known. Tests involving means and differences between means assume that the variance is known. Similarly, the convenient test of hypotheses involving the standard deviation assumes that the mean is known; a more laborious method is available when it is unknown. (This test for the standard deviation is presented in [3, 5, 14].)

Hypothesis involving means. Application of the basic equations for testing hypotheses concerning means as-sumes that observations are drawn from a normally distributed population. These equations, defining the parallel decision lines, are

$$\frac{u_1 - u_0}{\sigma^2} \Sigma X_i + \frac{u_0^2 - u_1^2}{2\sigma^2} n$$
$$= \log_e \frac{\beta}{1 - \alpha}, \quad \text{and} \quad [5a]$$

$$\frac{u_1 - u_0}{\sigma^2} \Sigma X_i + \frac{u_0^2 - u_1^2}{2\sigma^2} n$$
$$= \log_e \frac{1 - \beta}{\alpha} \quad [5b]$$

where u_0 and u_1 are the means stated in our two hypotheses, H_0 and H_1, respectively, X, is a single observation, σ^2 is the variance of X, n is the number of observations, and α and β are the specified probabilities of Type I and Type II errors. Since these formulas involve the variance, the experimenter must either have a good estimate of the variance from other data or must take an initial group of cases (n) of sufficient size to yield a

reasonably reliable estimate σ^2 of the variance. If the estimate is too small (or too large), he will be unwittingly increasing or decreasing the values of α and β and thus will make his test less (more) sensitive than he has planned to have it.

By solving formulas [5a] and [5b] for ΣX_i, we obtain the following equations:

$$\Sigma X_i = \log_e \frac{\beta}{1-\alpha} \frac{(\sigma^2)}{(u_1 - u_0)}$$
$$+ \frac{(u_1 + u_0)n}{2}, \quad \text{and} \quad [6a]$$

$$\Sigma X_i = \log_e \frac{1-\beta}{\alpha} \frac{(\sigma^2)}{(u_1 - u_0)}$$
$$+ \frac{(u_1 + u_0)n}{2}. \quad [6b]$$

These are the equations for the parallel lines in a graph representing the sequential plan for testing hypotheses involving means. To carry out the test, we plot on the graph the observed ΣX_i for each value of n. When the observed cumulative sum falls outside one of the lines, the appropriate decision is made.

Difference between means. A convenient sequential procedure is available for testing the significance of the difference between means of two equal-sized samples. In almost any instance where the sequential approach would be preferred to a conventional method based upon fixed N's, the use of equal-sized samples is not a serious restriction.

The procedure involves the use of difference scores between randomly selected pairs, or matched pairs if the groups are matched case for case. The sum of these difference scores is plotted on the figure specified by formulas [6a] and [6b], with u_0 set at zero and u_1 determined by theoretical

or practical considerations. Once again, the procedure assumes that the variance is known. The variance used here is, of course, the variance of the difference between means (for correlated or uncorrelated samples, whichever is appropriate).

Two-sided hypotheses involving means. All the preceding tests were one-sided; for example, the tests involving means specified just two values, u_0 and u_1. Even the test of mean difference required a single specific value for the alternative hypothesis. The procedure outlined below is designed to test whether the mean of a group deviates in *either* direction from some specified value. Thus instead of setting one value for the alternative hypothesis, we set two, which deviate the same amount but in opposite directions from the hypothesized central value, u_0. (For a full account, see 3.)

Let d be the deviation of each of the two means, specified under the alternative hypothesis, from the mean, u_0, specified in the null hypothesis. Let our critical values be

$$a = \log_e \frac{1-\beta}{\alpha}, \quad \text{and}$$

$$b = \log_e \frac{\beta}{1-\alpha}.$$

The formula required for the sequential test of significance is

$$G = \log_e \cosh \frac{d| \Sigma(X_i - u_0)|}{\sigma^2} - \frac{d^2 n}{2\sigma^2}, \quad [7]$$

where n is the number of observations (as previously), X_i is any observed score, σ^2 is the true variance of the population from which the observations are drawn, and $\cosh$ refers to the hyperbolic cosine. Tables of the hyperbolic cosines are accessible (e.g., **2**, Table XV; **3**, Table 5.2). If G is

greater than a, the null hypothesis is rejected in favor of the alternative hypothesis that the mean of the group differs from u_0 by the amount d. If G is more negative than b, we accept the null hypothesis that the mean of the group does not differ from u_0.

For most values, it is unnecessary to use the tables for the hyperbolic cosine because $\log_e \cosh z$ is asymptotic to $z - \log_e 2$ or $z - .693$. The equivalence is exact to two decimal places if $z \geqq 2.7$. Hence we may modify formula [7] as follows:

$$G' = \frac{d|\ \Sigma(X_i - u_0)|}{\sigma^2} - \frac{d^2}{2\sigma^2} n - .693. \qquad [8]$$

We set formula [8] equal first to a and then to b, and solve for $|\Sigma(X_i - u_0)|$:

$$|\Sigma(X_i - u_0)| = \frac{d}{2} n + \frac{\sigma^2(a + .693)}{d}, \text{ and } [9a]$$

$$|\Sigma(X_i - u_0)| = \frac{d}{2} n + \frac{\sigma^2(b + .693)}{d}. \qquad [9b]$$

These equations yield the parallel lines in Fig. 5, which has been set up for the following example: $u_0 = 10$, $\sigma^2 = 100$, $d = 6$, $\alpha = .10$, $\beta = .05$. Data for three sampling experiments are plotted on Fig. 5, the values being obtained by drawing random sequential samples from normally distributed populations with $u = 9$, 16, and 20 respectively.

An important consideration must be emphasized. Equations [9a] and [9b] were based upon equation [8], which is an adequate approximation to the precise equation [7] so long as $z \geqq 2.7$, where
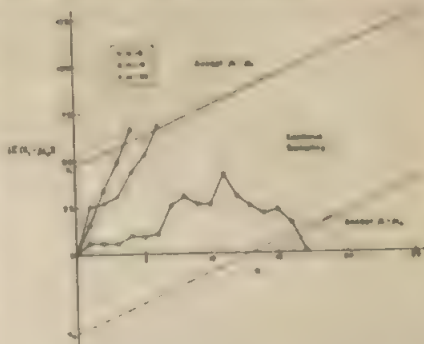
FIG. 5. A SEQUENTIAL PLAN FOR TESTING HYPOTHESIS ABOUT MEAN AGAINST TWO-SIDED ALTERNATIVE

$$z = \frac{d|\ \Sigma(X_i - u_0)|}{\sigma^2}.$$

For $z < 2.7$, equation [7] must be applied. It is sufficient to evaluate three or four values of G, corresponding to values $0 \leqq z < 2.7$, from equation [7]. These values provide points on the curved portion of the lower decision line of Fig. 5, and allow the necessary correction to become part of the graph of the sequential plan.

In Fig. 5, the sampling experiment for $u = 20$ gives a value of 70.32 at $n = 6$, allowing us to reject $H_0: u = 10$, in favor of $H_1: u = 10 \pm 6$. The sampling experiment for $u = 16$ allows the same decision at $n = 4$. For $u = 9$, it has required an n of 17 to yield a decision, in this case acceptance of H_0 (with probability of erroneous decision, $\beta = .05$).

The application of this method to a two-sided test of the difference between two means is clear-cut. Once again we take pairs of observations, one from each group, and analyze the differences for the n pairs. The value of u_0 is of course zero.

Hypotheses involving correlations. No convenient method is currently available for utilizing sequential

analysis in tests of hypotheses involving correlations. (It is possible that the labor involved in the computation of correlations may offset the advantages of sequential design.) While the product-moment correlation coefficient, r, is a mean, the mean cross product of the pairs of observations in standard score form, the sampling variance is a function of the population correlation and therefore would differ for two hypothesized correlation coefficients.

In many cases, a relationship can be measured by some statistic other than r. However, the nonparametric statistics, rho and tau, are unsatisfactory for sequential tests. For each, as for r, the variance is a function of the relationship in the population. Furthermore, the sampling distributions of rank coefficients for the nonnull case are complex; the approximation in the case of tau (9, chs. 4 and 5) seems too crude for use in sequential tests.

In this connection, Wolbers (17) has suggested a procedure for estimating correlation coefficients as data accumulate. In the method he suggests, each variate is dichotomized at the median. He provides a table for estimating the relationship from the number of cases in the accumulated sample and the number of cases falling on different sides of the two medians (e.g., above the median on X and below the median on Y). However, this method yields an estimation. It is not a sequential test of hypotheses.

A general test of hypotheses involving double dichotomies. A sequential test is available for determining which of two groups has the higher proportion of some specified condition. This test applies to the general case of testing differences in proportions of "successes" in two samples. It could have been used for an experiment on the problem involved in our original example, that of comparing the proportions of successes produced by two therapeutic techniques. Patients would be assigned randomly to two groups, one for each type of therapy, and the proportions of successes in each would be ascertained empirically.

Such a test, for difference between proportions, should be contrasted with that test applied in the earlier example to test the significance of the deviation of a sample proportion from an a priori value. Since the test model for double dichotomies is more complex than the methods previously discussed, the reader is referred to the original source for details (3, sec. 3).

GENERAL REMARKS

The selection of the alternative hypothesis. The alternative hypothesis tested in a sequential test specifies a parameter other than that given by the null hypothesis. How does the experimenter select this value? We may distinguish two general situations. In the first situation the experimenter can specify a value for H_1 on one of three particular bases. (*a*) He may be testing a theoretical estimation of the parameter. Rarely would this basis be available to the psychologist, since current theory does not seem sufficiently precise. (*b*) He can choose a value on the basis of previous empirical findings. (*c*) He can decide on practical grounds that the parameter must be at or beyond the specified value because lesser values would make no practical difference. Some of the earliest applications of sequential analysis were to problems of sampling inspection where the parameter presumably was chosen on such grounds.

But another general situation may exist where none of the above bases

for decision is available or appropriate. The experimenter may wish to use a sequential test for the sole purpose of economy of effort. What rationale can he follow? For these circumstances, we suggest a plan based upon a rough approximation to the largest number of observations that the experimenter is willing to use. Having chosen a trial H_1 such that for selected values of α and β the OC curve is satisfactory, the experimenter can compare the highest point on the ASN curve with his preferred maximum N. If the maximum ASN exceeds the preferred maximum N, a second value of H_1, more distant from H_0 than the first, may be chosen and new OC and ASN curves drawn.

The chance that the actual sample size required to make a decision will exceed the maximum ASN is small, definitely below .5. Usually one of the hypotheses will be accepted long before the sample size reaches the maximum ASN, and often when the number of observations is less than half that size.

It is obvious that the closer together the values postulated in H_0 and H_1, the larger the ASN (cf. Fig. 4). As in conventional statistics, small effects are more difficult to isolate than large ones. On the other hand, if the difference between the two hypothesized values is unreasonably large, the experimenter gains little by deciding to accept one hypothesis in preference to the other.

Even when an alternative hypothesis can be determined on a priori or practical grounds, it is advisable to consider the OC and ASN curves in planning a sequential test. Only in this way can an experimenter be completely assured that his design is one which actually will provide the test he desires at an acceptable cost (moderate ASN).

Sequential tests vs. conventional tests. Does a sequential test have any advantages when the selection of a second hypothesis is done on an essentially arbitrary basis? The chief advantage is economy. Sequential analysis produces an *average* saving in sample size of about one-third to one-half, when compared with the most powerful current statistical tests. However, the experimenter using sequential analysis will use *more* cases about 25 per cent of the times he uses this test if $\alpha = \beta = .05$. (Rarely will he have to use twice as many cases.) If $\alpha = \beta = .01$, the probability is about .9 that he will effect a saving (cf. 14, p. 57 and p. 60).

This advantage can be seen intuitively in our first example. Here, the two hypotheses stipulate the proportion of successes at .60 and .75 respectively. If our first eight randomly selected cases had included no failure case, it would be very unlikely that the observed sample p subsequently would drop to a value so low as to lead to acceptance of $H_0 : \pi = .60$. On the other hand, if six of the eight were failures, it would be unlikely for subsequent sampling to produce a p so high as to lead to acceptance of $H_1 : \pi = .75$. Such extreme results are uncommon. However, sets of observations showing some tendency toward an extreme occur sufficiently often to make the method economical. Sequential analysis takes advantage of the fact that some possible sets of n observations make improbable (or impossible) certain obtained values for $n+1$, $n+3$, or even $3n$ observations.

An evaluation of the sequential design. The uncritical general use of sequential analysis obviously is not recommended. It is a design which can have advantages when one or more of the following conditions

holds: (a) The problem involves the choice between two possible parameter values which can be specified on a priori but not arbitrary grounds (see "The Selection of the Alternative Hypothesis," above). The null hypothesis will usually be one of the two. (b) The data are such that the cost per datum is high and economy is desired. (c) The total amount of data is not fixed. In this situation, the experimenter takes a calculated risk. We have seen above that in a small proportion of applications of sequential tests, he will require more cases than he would if he used a more conventional statistic. However, the odds are great that he will economize on his data collection.

As Wallis (7, ch. 17) points out, in the nonsequential test, N is fixed, and with one risk specified (α), the second risk (β) is minimized. In the sequential test, we specify both risks (α and β) and minimize the sample size.

In a situation where we are able to specify both α and β, but where it is uneconomical to collect observations piecemeal, it may be preferable to estimate the desirable sample size and proceed with a conventional study, collecting and analyzing all data at the same time (cf. 5, ch. 14, especially pp. 212–213 and pp. 219–220; also 6, pp. 153–157). For this latter situation, the total effort required to reach a decision can be estimated in advance. In a sequential test, such an estimate will necessarily be only a rough approximation.

Summary

This paper seeks to introduce sequential analysis to psychological researchers as a statistical design that has considerable promise for work in areas where the cost per observation is high.

The basic design involves the testing of alternative hypotheses, one of which is usually a null hypothesis. The experimenter sets up three alternative decisions, one of which is accepted after each set of data is analyzed: (a) he may accept the first hypothesis; (b) he may accept the second hypothesis; (c) he may accept neither and take another observation. If he makes the third decision, he makes another observation (or set of observations) and then repeats the analysis. This procedure is followed until he accepts one hypothesis or the other.

In its more convenient forms, the sequential test assumes knowledge of the population variance or (in tests of variance) knowledge of the population mean. The applications of the statistical design to problems involving proportions, means, and mean differences are illustrated. Methods are given for sketching operating characteristic and average sample number curves for sequential tests. The use of these curves is exemplified and their value is emphasized.

REFERENCES

1. ANASTASI, ANNE. An empirical study of the applicability of sequential analysis to item selection. *Educ. psychol. Measmt*, 1953, 13, 3–13.
2. BURINGTON, R. S. *Handbook of mathematical tables and formulas.* (3rd Ed.) Sandusky, Ohio: Handbook Publishers, 1949.
3. COLUMBIA UNIVERSITY, STATISTICAL RESEARCH GROUP. *Sequential analysis of statistical data: applications.* (Rev. Ed.) New York: Columbia Univer. Press, 1945.
4. COWDEN, D. J. An application of sequential sampling to testing students. *J. Amer. statist. Ass.*, 1946, 41, 547–556.
5. DIXON, W. J., & MASSEY, F. J., JR. *An introduction to statistical analysis.* New York: McGraw-Hill, 1951.
6. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

7. EISENHART, C., HASTAY, M. W., & WALLIS, W. A. (Eds.) *Techniques of statistical analysis.* New York: Mc-Graw-Hill, 1947.

8. JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

9. KENDALL, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.

10. KIMBALL, A. W. Sequential sampling plans for use in psychological test work. *Psychometrika*, 1950, 15, 1–15.

11. MOOD, A. McF. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

12. MOONAN, W. J. Some empirical aspects of the sequential analysis technique as applied to an achievement examination. *J. exp. Educ.*, 1950, 18, 195–207.

13. SCHMID, J., JR. Sequential analysis of test items. *J. exp. Educ.*, 1952, 20, 261–264.

14. WALD, A. *Sequential analysis.* New York: Wiley, 1947.

15. WALKER, HELEN M. Item selection by sequential sampling. *Teachers Coll. Rec.*, 1949, 50, 404–409.

16. WALKER, HELEN M., & COHEN, S. *Probability tables for item analysis by means of sequential sampling.* (Prelim. Ed.) New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1949.

17. WOLBERS, H. L., JR. Estimation of correlational relationships using the techniques of sequential analysis. *Amer. Psychologist*, 1950, 5, 463. (Abstract)

SPEED OF LEARNING AND AMOUNT RETAINED: A CONSIDERATION OF METHODOLOGY

BENTON J. UNDERWOOD[1]

Northwestern University[2]

A common generalization found in textbooks is that the fast learner retains more than the slow learner. McGeoch (**6**), apparently on the basis of Gillette's work (**1**) and on the basis of his general conception of learning and forgetting processes, wrote: "This high positive relation between individual scores in learning and retention is to be expected, of course, from the fact that these are continuous processes" (**6**, p. 388). Although not explicitly stated in terms of rate of learning and forgetting of individuals, the present writer has also assumed the validity of the general principle that speed of learning and amount retained are directly related (**8**, p. 510). Kingsley (**4**, p. 469), Munn (**7**, p. 213), and Hilgard (**2**, p. 10) have also accepted the generalization.

The present paper has two purposes. First, we will show that the methods used to obtain the data from which this principle was derived are inappropriate. Secondly, we will show that when suitable methodology is applied to the problem, there is no difference in rate of forgetting of fast and slow learners.

METHODS USED BY PREVIOUS INVESTIGATORS

The last extensive work on the problem of the relation between speed of learning and amount retained was Gillette's monograph (**1**) published in 1936. It is highly probable that this monograph is responsible for the general conclusions appearing in textbooks as noted above. In this monograph Gillette critically summarizes the methods and results of all previous investigations, going back as far as Ebbinghaus. Two methods of investigation had been used in these previous investigations. Gillette's criticisms of these methods may be noted briefly.

Method of equal amount learned. By this method all *S*s learn a task to a given criterion of proficiency, and after a rest interval retention measurements are obtained. The most commonly used statistic is the correlation coefficient between a measure of time to learn to the standard criterion and the recall or relearning scores. Gillette points out that this technique favors retention of the slow learner since certain of the items or parts of the task will be much overlearned and therefore might be expected to enhance retention. This would tend to reduce differences in retention between fast and slow learners. Indeed, correlations near zero between time to learn and amount recalled are interpreted by Gillette as actually indicating the slow learners have forgotten more rapidly than fast learners. Nevertheless, Gillette rightfully rejects this method as a means of reaching a definitive conclusion on the issue.

Method of equal opportunity to learn. By this method all *S*s are allowed the same amount of time to study the

material before the retention interval. Again, Gillette notes that this method is quite inappropriate because the fast learners will learn more than the slow learners and therefore may be expected to have higher retention simply because they have learned more. In general, the studies she reviews show fairly high correlations between amount learned and amount retained, and her own data resulting from this technique likewise yield high correlations. Actually, of course, using this technique, fast *S*s could show more rapid forgetting than slow *S*s, and the high correlations would still obtain if there were wide differences in amount originally learned. This method, therefore, must be rejected as a means of investigating the problem.

Method of adjusted learning. Gillette's solution to the methodological problem was to use the adjusted-learning technique. By this technique items in a list are "dropped out" after being learned to a minimum criterion (such as one correct anticipation). Thus, all items of a list are given the same number of "reinforcements." And, while slow learners would take more time to obtain these reinforcements than would fast learners, Gillette assumed that the strength of associations at the end of learning was the same for fast and for slow *S*s.

Gillette's conclusion from her experiment using the method of adjusted learning is as follows: "The slow learner when given sufficient time to learn the *same amount* as the fast learner, but not allowed to *overlearn* the material, is not able to retain as much as the fast learner" (1, p. 50). Her final conclusion, after using all three methods, was that there are very clear indications that the fast learner retains better than the slow learner.

INADEQUACY OF THE METHOD OF ADJUSTED LEARNING

The inadequacy of the method used by Gillette can best be understood if a brief résumé is given of certain calculations made on data which were available to us. As a result of a series of investigations on rote learning which we have published elsewhere in connection with other problems (9, 10, 11, 12), we had a large amount of data which could be analyzed in an attempt to obtain more information on the relationship between speed of learning and amount retained. These experiments involved both paired-associate and serial learning, and both nonsense syllables and adjectives. The lists were always learned to a criterion of one perfect recitation with recall and relearning occurring after 24 hr.

Using these data, we first obtained correlations between trials to learn and recall, and between trials to learn and relearn. Since our *S*s all learned to the same criterion, the method is the one referred to above as the method of equal amount learned. In general conformance with previous findings summarized by Gillette, correlations between time to learn and amount recalled were all near zero. Correlations between time to learn and time to relearn were high and positive. Viewed by this method, then, our results are quite in accord with previous findings.

Our next step was to analyze the data in such a fashion that we would in effect be using the adjusted-learning technique. We divided a group of *S*s for a given experiment into two equal subgroups based on number of trials to learn, thus giving a group of slow learners and a group of fast learners. We then counted the number of correct anticipations made by each *S* for each item in learning a list

to one perfect trial. Each correct anticipation we call a reinforcement simply because the greater the number of correct anticipations the more resistant is an item to forgetting. All items having the same number of reinforcements were pooled, using separate pools for slow and fast Ss. Thus, we might have 200 items which had been given five reinforcements while being learned by slow Ss and 225 items which had been given five reinforcements while being learned by fast Ss. The final step was to determine the percentage correct at recall for each frequency of reinforcement for fast and slow Ss. This is, in effect, comparable to Gillette's adjusted-learning technique.

Results of such comparisons between fast and slow Ss for several experiments showed consistently that for items equated for number of reinforcements during learning, retention was consistently superior for fast learners. This was true regardless of the actual number of reinforcements obtained during original learning. That is, fast Ss were superior in recall for items having only a few reinforcements and for items having many reinforcements. Thus, this analysis gave strong empirical confirmation to Gillette's findings.

The assumption on which acceptance of Gillette's method (and of our modification of it) rests is that the making of a correct anticipation (a reinforcement) by a slow S results in the same associative strength as a reinforcement for a fast S. If this assumption is not valid, comparisons of retention of fast and slow learners are as misleading by this method as by the other methods thus far discussed. For, if a reinforcement adds more strength to an association being acquired by a fast S than it does to one for a slow S, it means that degree of learning of the material is not equal before introduction of the retention interval. Comparisons of retention, therefore, are distorted in the same fashion as with the other two methods discussed above.

In order to make comparisons of retention of fast and slow learners, it is absolutely essential that the response strengths of the items for fast and slow learners be equated at the end of learning—at the start of the retention interval. The method of adjusted learning and our variation on this method do not assure that this is the case. Indeed, it is a very reasonable hypothesis that the essential difference between fast and slow learners is that a reinforcement *does* result in more associative strength for a fast S than for a slow S. Thus, the better retention of fast Ss as found by Gillette and also by our method may be entirely a function of the unequal degree of learning before the retention interval was introduced.

TECHNIQUE FOR EQUATING DEGREE OF LEARNING FOR FAST AND SLOW SUBJECTS

The solution to the problem results from what we call a successive probability analysis of learning. The data we will use to demonstrate the method come from five experiments on the learning and retention of paired nonsense syllable lists (10). Each experiment employed 36 Ss. We have divided these into two groups of 18 each based on number of trials to learn. Each S learned and recalled three lists, the conditions for learning differing only in terms of intertrial interval. These intervals (4, 30, and 60 sec.) produced no difference in learning, and they do not interact with S's ability level. Therefore, we may treat them as three equivalent learning tasks for each S.

Since there were five experiments, and since we get 18 fast Ss and 18 slow Ss from each experiment, we have a total of 90 fast and 90 slow Ss, each having learned and recalled three lists. Effectively, then, we have 270 lists for each group, and since each list had 10 pairs, we have 2,700 items for each group.

The successive probability analysis consists of determining the growth of the associative function for each item. We first observed when an item was correctly anticipated initially. We then noted whether the response was anticipated correctly or incorrectly on the immediately following trial. When an item had been anticipated correctly twice (whether on successive trials or not), we noted whether it was correct or incorrect on the next trial. This continued until the item was analyzed for the entire course of learning. As an illustration, suppose that an item was first given correctly on trial 6. On trial 7 it was also given correctly, on trial 8 incorrectly, trial 9 correctly, trial 10 correctly, with S also getting all other items correct on trial 10. Our entries would be: after 1 reinforcement, it was correct on next trial; after 2 reinforcements it was incorrect on next trial; and after three reinforcements, correct.

When such an analysis is made for each item for a large number of Ss, and when these data are pooled, we are in a position to make exact statements of the probability of getting an item correct on a succeeding trial when this item has been correctly anticipated on previous trials once, twice, three times, etc.

As can be seen, the labor involved in making this successive probability analysis is great. To reduce the labor somewhat we eliminated some items which were very easy to learn and

hence had been given a great many reinforcements. Working first with data for slow Ss, we eliminated all items which had been correctly anticipated more than 31 times by S in learning a list. Out of the 2,700 items, 289 were eliminated by this criterion. We then made a distribution of number of reinforcements given the 2,700 items by the fast Ss. On the assumption that easy items for slow Ss would also be the easy items for fast Ss, we eliminated approximately 289 of the easiest items for the fast Ss. The actual number was 322 since we did not want to eliminate a portion of a category, as would have been necessary if we had eliminated exactly 289 items. For these fast Ss, the elimination of 322 items left only items which had been reinforced less than 18 times.

The successive probability curves for fast and slow Ss are plotted in Fig. 1. The number of observations decrease as the number of reinforcements increase. For the fast Ss a total of 18,379 observations are involved, with 2,372 for the case of one reinforcement and 54 for the 17 reinforce-
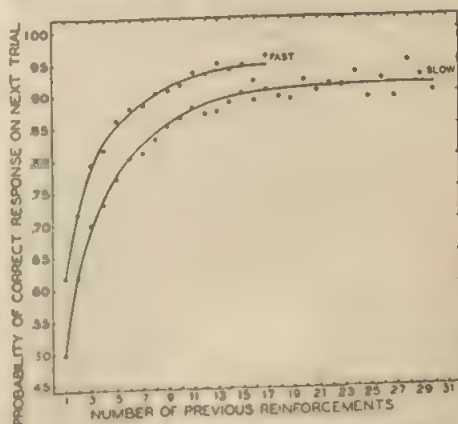

FIG. 1. THE RELATIONSHIP BETWEEN NUMBER OF REINFORCEMENTS DURING LEARNING AND THE PROBABILITY OF CORRECT RESPONSE ON NEXT TRIAL FOR FAST AND SLOW LEARNERS

ments. For the slow Ss 31,560 observations were totalled, with 2,403 for one reinforcement and 28 for 30 reinforcements. The amount of labor involved in compiling the data for the curves has some compensation in the smoothness of the functions.

From Fig. 1 it is clear that a reinforcement on a given trial (or series of trials) does not result in equal probabilities that that response is correct on the next trial for fast and slow Ss. Following one reinforcement, for example, the probability is .62 for a fast S that the item will be correctly anticipated on the next trial. For the slow Ss the value is .50. Such differences occur throughout the entire range of reinforcements plotted. It is thus apparent that studies which have "equated" fast and slow learners by equating number of reinforcements have not equated for associative strength. Therefore, our next step is to equate associative strength at the end of learning so that comparable retention measurements may be made for fast and slow Ss.

It can be seen from Fig. 1 that the equation of associative strength could be made graphically. Thus, we see that six reinforcements for slow Ss result in about the same associative strength during learning as do three reinforcements for fast Ss. Other comparable values could be determined and two base lines plotted, one for fast Ss and one for slow Ss, in such a fashion that associative strength at the end of learning is equal for the two groups at any point along the base lines. Such adjustments would allow us to say that if these Ss had been measured on an immediately succeeding trial following the end of learning, the associative strength for fast and slow Ss would have been the same. Knowing this, we can then make observations of recall with con-

fidence that differences in retention are a function of differences in the forgetting of fast and slow Ss, and not a function of differences in response strength at the end of original learning.

In general, the above procedure has been followed. The critical step, of course, is to equate response strengths along the base line for fast and slow Ss. To do this we first gave mathematical expression to the curves of Fig. 1. After trying several types of curves, it was found that the exponential-type function gave the best fits. Essentially the formula is that of Hull (3, p. 119), with the exception that we have not forced the curves to pass through the origin. The two formulas are:

Fast Ss: $H = 95.6 - 35.20e^{-.1081N}$

Slow Ss: $H = 92.5 - 40.91e^{-.0801N}$.

In these expressions H is associative strength. The first values on the right-hand side (95.6 and 92.5) are asymptotes estimated from Fig. 1. The notation e is a constant given the value of 10, and N is the number of reinforcements. These formulas fit the data well except for the first point on each curve. Details of the procedure used in deriving and working with these curves may be found in Lewis (5, pp. 61–62).

To construct exactly comparable base lines using the above formulas we used the curve for fast Ss as a reference curve. The H values for each successive N for fast Ss were substituted into the formula for slow Ss and the equation solved for N. This means that we could determine the N value for slow Ss which gave the same H value for a given N value for fast Ss. Thus, for each H value we have two N values, one for each group. The two N values for successive H values

are plotted on coordinated base lines in Fig. 2. At any point along these base lines we know that the probability that a response will be correct on an immediately succeeding trial is the same for both groups. In Fig. 2 the highest N value for fast Ss is 10, this being almost equivalent in associative strength to N of 30 for slow Ss.

The final step is to plot the proportion of items recalled after 24 hr. for each group for each level of associative strength. When this is done (Fig. 2) we see that no consistent difference exists between the two curves. We conclude that when associative strength at the end of learning is equivalent for slow and fast Ss, no difference in recall after 24 hr. may be expected. This holds true for a wide range of associative strengths. Because forgetting is rather great in both groups, and because it does not vary between groups as a function of associative strength, we would not expect differences to emerge with longer retention intervals.

OTHER IMPLICATIONS

The results of the above analyses lead to a strong temptation to assert that the critical difference between fast and slow Ss is that the associative strength resulting from a reinforcement is less for slow than for fast learners. In a descriptive sense this is true, as indicated by Fig. 1. However, in the theoretical sense some reservation is necessary. Between successive presentations of a given item during learning, a maximum of about 1.5 min. elapsed. This interval might be as low as a few seconds depending upon the location of the item in successive presentations of the list and upon the intertrial interval used. It might be argued that if there is no difference in forgetting between fast and slow Ss
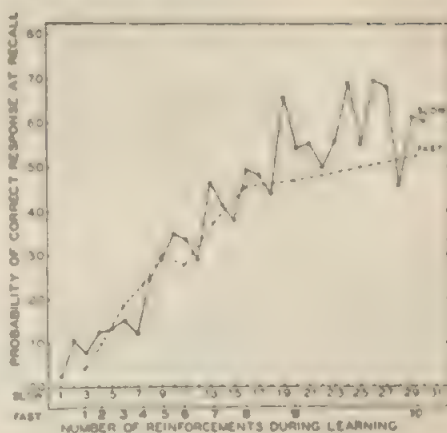
FIG. 2. RETENTION OF FAST AND SLOW LEARNERS AFTER 24 HR. WHEN ASSOCIATIONS ARE EQUATED FOR STRENGTH AT THE END OF LEARNING

over 24 hr., it is highly unlikely that there are differences in forgetting over these very short intervals between successive presentations of the same item during learning. Nevertheless, it is possible that a reinforcement for a fast and a slow S adds the same associative strength, but that in the context of learning a group of items (as contrasted with the relative rest over 24 hr.) more forgetting might take place for an item for slow Ss than for fast Ss over these short intervals between trials. Thus, it must not inevitably follow that reinforcements add different associative strengths at the time they occur for fast and slow Ss; we believe they do add or produce different amounts of associative strength but our data are not conclusive on this matter. We can conclusively say only that when probabilities of response for successive trials during learning are equated for fast and slow Ss, no difference in forgetting occurred.

The relationships between other variables and forgetting may, if ana-

lyzed by the present technique, be different from those commonly believed to be true. However, such analyses may be expected to be fruitful only if differences produced by the variable during *learning* are fairly large. For example, it is generally believed that the more meaningful the material the less rapid the forgetting. Meaningfulness *does* produce large differences in rate of learning. However, it is quite possible that if materials or different degrees of meaningfulness were equated for associative strength at the end of learning, no differences in rate of forgetting would be demonstrable. It goes without saying that if such findings do occur we will need to shift our ideas when constructing theories of forgetting.

SUMMARY

A review of previous methods used to study the relationship between rate of learning and rate of forgetting showed that none was adequate to the problem. In no case was it established that associative strength before the retention interval was equal for *S*s learning at different rates.

A method for equating associative strength at the termination of learning was described and applied to data obtained in learning and recalling paired nonsense syllables. With such equality established it was shown that no difference in forgetting occurred over 24 hr.; the recall of slow *S*s was as good as that of fast *S*s.

The method may be useful for studying the influence of other variables on retention.

REFERENCES

1. GILLETTE, ANNETTE L. Learning and retention: a comparison of three experimental procedures. *Arch. Psychol.*, 1936, **28**, No. 198.
2. HILGARD, E. R. *Introduction to psychology.* New York: Harcourt, Brace, 1953.
3. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. KINGSLEY, H. L. *The nature and conditions of learning.* New York: Prentice-Hall, 1946.
5. LEWIS, D. *Quantitative methods in psychology.* Iowa City, Ia.: Gordon Bookshop, 1948.
6. McGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
7. MUNN, N. L. *Psychology.* (2nd Ed.) New York: Houghton Mifflin, 1951.
8. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
9. UNDERWOOD, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
10. UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.
11. UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45** 143–149.
12. UNDERWOOD, B. J. Studies of distributed practice: X. The influence of intralist similarity on learning and retention of serial adjectives. *J. exp. Psychol.*, 1953, **45**, 253–259.

SPECIAL REVIEW

SOME RECENT BOOKS ON PERSONALITY

DAN L. ADLER
San Francisco State College

BLUM, GERALD S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953. Pp. xviii+219. $3.75.

EYSENCK, H. J. *The scientific study of personality.* New York: Macmillan, 1952. Pp. xiii+320. $4.50.

EYSENCK, H. J. *The structure of human personality.* New York: Wiley, 1953. Pp. xix+348. $5.75.

HARROWER, MOLLY. *Appraising personality.* New York: Norton, 1952. Pp. xvii+197. $4.00.

JERSILD, ARTHUR T. *In search of self.* New York: Columbia Univer. Teachers Coll., 1952, Pp. xii+147. $2.75.

NOTCUTT, BERNARD. *The psychology of personality.* New York: Philosophical Library, 1953. Pp. 259. $4.75.

NUTTIN, JOSEPH. *Psychoanalysis and personality.* New York: Sheed & Ward, 1953. Pp. xiv+310. $4.00.

PATTY, WILLIAM L., & JOHNSON, LOUISE SNYDER. *Personality and adjustment.* New York: McGraw-Hill, 1953. Pp. viii+403. $4.75.

SPINLEY, B. M. *The deprived and the privileged.* London: Routledge & Kegan Paul, 1953. Pp. vii+208. $4.00.

VERNON, PHILIP E. *Personality tests and assessments.* London: Methuen, 1953. Pp. xi+220. 18s.

Introduction: A Frame of Reference

If, as Smith (4) maintains, social psychology attained its rapid growth rate and transformation during recent postwar years, then it can be said that the study of personality lost its momentum during approximately this same period. Indeed, the latter seems now to have come to another node of its epicycloidal career, although not quite back to its starting point. A look at this last cycle will give some perspective to the reviews that follow.

The decade preceding the war was characterized by an exhilarating drive to explore the whys and wherefores of behavior. The air was alive with dynamic theories, "models," and experimental thrusts into the nature of motivated action. Genotypical soul mates like Tolman's "sowbug" and Lewin's "baked potato" vied with Hullian *habs* under the watchful eyes of mathematicians, physiologists, and logical positivists. What quarreling existed was intellectually sound and was conducted, in the main, under Marquis of Queensbury rules of scientific inquiry. This was a period when evolutions inherent in the development of psychology reached apogee. It was then, for example, that atomism seemed finally to be unseated by molar interdependence, when instruments became less important than behaviors, and when classification and empirical law were supplanted by theory. It was then that psychology embraced the hypothetico-deductive method of science and began truly to shed its metaphysical trappings.

It is difficult to recognize these prewar antecedents in postwar psychology, particularly in the area of personality. There is, of course, no less concern with the causes and courses

283

of behavior, but the method of explication has undergone radical change. A major divergence has become explicit. There is, on the one hand, an experimental group involved in the mechanics of proof, operating under improved ground rules of statistical design and rigor. On the other hand, there has emerged an equally vigorous clinically oriented discipline that, for the most part, operates with different assumptions and for whom the nature of proof basically seems to be "self-evidence."

For rather different reasons, both groups appear to be contributing to the delinquency of psychology by disinheriting the scientific bent so laboriously acquired during prewar years. The experimentalists' efforts are now all too often vitiated by absorption in details and specious variables. Unlike their predecessors, they seem to have narrowed rather than broadened the breadth of their conceptualization and pursuits. There is a danger that the hypothetico-deductive method will be lost to them for lack of a theoretical vehicle to carry it. The clinical group presents the problem differently. They have not forsaken scientific method—principally because they have never adopted it. Although their explanatory systems have rarely lacked breadth,they have continuously lacked evidence. Where the experimentalist is inclined to recast his system on the basis of a single contraindication, the clinician is apt to "prove" his system by a single bit of affirming evidence.

This picture of contrast between the pre- and postwar psychology of personality is painted, of course, without qualifying "shades of gray." It is not so much a lament for personality psychology as it is a notation of the *status quo* from which we will ultimately move. Thus, the current literature contains not only manifest evidence of the present position as postulated, but hints as well at some lines of future development.

Of the ten books included in this review, six come from other than American psychologists. The four American contributions may be classified basically as compendia of what we know, or think we know, about personality. One is a text (Patty and Johnson), one is a compilation of theories (Blum), and two are "practical guides" for physicians and teachers respectively (Harrower, Jersild).

The European contributors afford some relief to this uniformity, although three of these books are also compilations (Eysenck [1953], Vernon, Notcutt). The remaining three are of a different stripe altogether. One is Eysenck's (1952) recapitulation and development of a factorial theory of personality made known to Americans in 1947 (2). Nuttin's book is a declared attempt to reinterpret Freudian doctrine in what the author calls a "nonpathological" sense. Spinley's volume is a research report which differentiates the perceptions and personalities of individuals belonging to two English subcultures.

Eysenck—*Scientific Study of Personality*

Of all the books considered, this is at once the most provocative and the most provoking. It combines the virtue of sophisticated deduction and the vice of blatant bias. Although occasionally it carps and quibbles, the unfolding of its theme tantalizes to the very end.

Eysenck sets the stage in the first chapter with a review of the essentials of scientific method, specifically emphasizing antinominalism, conceptualization, and testability. Then,

considering the developmental stages of science, he consigns psychology to a bottom rung of this developmental ladder. Our need, according to Eysenck, is to establish a new and valid taxonomy of personality. The only method that will enable us to do do so now is factor analysis (p. 42). But not any factorial method suffices. Eschewing, for the noncognitive field, the methods of both Thurstone and Cattell, Eysenck presents "Criterion Analysis" as an alternative procedure, which purportedly combines the effectiveness of factor analysis and the hypothetico-deductive method.

Criterion analysis begins with the selection of criterion groups (e.g., normal and neurotic groups) and the administration of what might prove to be discriminating tests. Tetrachoric or biserial correlations with the criterion groups are obtained (criterion correlations) and test intercorrelations determined. Factor coefficients (saturations) are derived from the latter, and their proportionality to the criterion coefficients is examined. The presence of this proportionality is taken as evidence that the criterion factor (e.g., neuroticism) has been isolated *and* identified.

In Chapter 4, Eysenck reports the first full-scale use of this method. Employing an axis rotation which maximizes the correlation between factor saturation and criterion indices, he makes an attempt to identify the factor of neuroticism and its subcategories. It is here that one begins to raise questions about the author's presentation.

1. Having administered a battery of 76 tests, he selects 28 for factorial analysis. The basis of this choice is not made clear. Presumably, the number of subjects involved does not determine the choice since in three tests $N = 38$, and in the remainder $N = 96$. Nor does the choice seem to depend on the significance of the correlation ratios (pp. 127–150). One wonders what principle of elimination was used.

2. The selection of criterion groups is based upon psychiatric ratings with which Eysenck repeatedly finds fault until a positive attitude is useful to his argument. Thus " . . . psychiatric diagnosis is of doubtful validity and low reliability" (p. 33). Later, after a discussion of inconsistent inductee rejections on neuropsychiatric grounds, he says: "We are left with the unreliability of psychiatric assessment as the most probable cause" (p. 91). But, he then maintains, " . . . as we have shown in the preceding section, the validity of psychiatric ratings for future breakdown has been established; it would seem to follow that methods correlating highly with psychiatric assessment would also show a certain degree of validity" (p. 99). Practically all of the experiments which bolster his arguments for a neuroticism factor depend upon psychiatric diagnoses or ratings of severity of neurotic disorders (e.g., pp. 106 f.).

3. A similar tendency on the part of the author to favor expedient interpretations is exemplified by his variable interpretation of correlation coefficients. Thus, a correlation of $-.21$ between intelligence and neurotic tendencies is considered to be a "very small one" (p. 116) and it is interpreted as indicating no close relationship (p. 149). Nevertheless, a series of correlations between level of aspiration tests and examination scores "gave results in the expected direction," albeit they are $-.20$ and $-.11$ in one case, and $-.15$ and $-.10$ in another (p. 266).

There is little doubt that psychologists will follow Eysenck's logical reasoning about neuroticism, despite the lapses indicated. But the step which follows will separate the environmentalist sheep from the hereditarian goats. Employing Holzinger's h^2 as an estimate of the contribution of heredity to total variance, Eysenck concludes that neuroticism as a factor is a biological unit which is inherited as a whole (p. 187). The crux of his argument lies in the differences between identical and fraternal twins on tests which Eysenck employs to differentiate normals from neurotics. Although mean test scores are the same for the two types of twins, pairs of identical

twins are characterized by higher intracorrelations, and the group of identicals as a whole is characterized by greater variance.

One possible source of irritation to the reader should be mentioned here. Eysenck seems to have the happy faculty of anticipating objections to his theses, and the unhappy one of meeting these objections with a combination of affirmation and partial recant. The assumptions underlying the use of h^2 are probably not met, he says (p. 189), but they may err in the direction of *minimizing* the *true* contribution of heredity. On the same page, however, he points out the possibility of obtaining a *lower* h^2 value in another culture if the environmental factors of that culture were more pronounced. Presumably then, there might be a culture in which it could be "proved" that the total variance attributable to heredity is negligible.

Eysenck rejects the notion, attributed to Freud, that psychoticism is a part of a normal-neurotic-psychotic continuum. Although by criterion analysis he distinguishes between "psychopath" and "hysteria-anxiety" types of neuroticism, he finds little difference between manic-depressive and schizophrenic types of psychoticism. The reader may be inclined to wonder if the basic premises of criterion analysis are sound. Quite apart from Eysenck's conclusions, it is probably no more difficult to distinguish psychotic subtypes than it is neurotic subtypes. On a priori grounds one would expect psychiatric judgments to be at least equally valid in making these distinctions. Yet having employed the distinctions to create criterion groups, we are required to believe that the differences we "see" are not valid because test behaviors of the groups do not satisfy a statistical notion. In this instance, at least, we might conclude that criterion analysis is better used to discover what is not known than to corroborate what is.

Eysenck's book is an important one because of its generally convincing logic and its freshness of approach. These virtues may well operate to make it a new choice point in psychological history. Such importance makes it imperative to evaluate this factorial method in the light of our developmental frame of reference. Apart from the specific criticisms already made, two major points should be emphasized:

1. Criterion analysis presumably has the advantages of being a hypothetico-deductive process. Careful examination indicates, however, that the hypotheses are mainly validated internally, i.e., within the confines of the method itself. (See particularly the first full-scale application of the method described in Chapter 4.) The few exceptions in Chapter 7 notwithstanding, it fails to comply with the rigorous requirement of independent external validations.

2. Criterion analysis is admittedly elementaristic, and its sponsor is openly opposed to organismic interpretations of behavior or personality (pp. 277–285). Citing hand-picked experiments, he refutes the meaningfulness of gestalt qualities and substitutes for them scalar values on orthogonal dimensions. Nowhere is it clear how this index description of personality can help determine the dynamics—the causes of change—of behavior. Nowhere is it clear how measured body sway—a reliable index of "neuroticism"—will predict today's equanimity and tomorrow's moodiness. Eysenck himself cites the difficulties inherent in standardizing his factor-loaded tests:

" . . . whether the tester is a pretty young girl, a domineering male, or a mature, sympathetic woman—all these influences may, and in some cases do, affect a person's scores. They are difficult to control, particularly as the same stimulus—a pretty young girl as tester, for instance—may mean quite different things and arouse quite different emotions, in subjects differing in age, sex, and marital happiness" (p. 294).
He might well have added here that the same stimulus may be perceived differently by the

same subject from day to day or minute to minute. Not only the test, but the whole theoretical structure is thereby placed in jeopardy.

Eysenck—*The Structure of Human Personality*

In this latest book, we find neither essential changes nor further developments of Eysenck's theoretical position. It is useful primarily as a source book for those who would examine personality by the criterion of factor analysis. As such, it contains excellent summaries of factorial studies classified under the following chapter headings: "Analysis of Ratings," "Analysis of Questionnaires and Inventories," "Analysis of Objective Behavior Tests," "Analysis of Physique," "Analysis of Physiological Measures," "Analysis of Interests and Attitudes," "Analysis of Correlations between Persons," and "Analysis of 'Trait' Measurements."

Each chapter comprises a historical survey and a critical analysis of the area under consideration from the standpoint of its experimental worth, conceptualization, and the factor methods employed. For those who have read *The Scientific Study of Personality*, it goes without saying that the critique is made through no ordinary orthogonal eyes, but through maximally rotated Eysenckian ones. In this regard, it is regrettable that Eysenck does not maintain throughout the sense of humor which shows up so well in the Preface: " . . . it might be wise to follow psychoanalytic practice and lay it down that criticism of factor analysis should be confined to those who had themselves been factor analysed!" (p. xiv). We are treated, instead, to a steady display of the argumentative ingenuity and the perspicacity that are so characteristic of the author.

Although the book is intended in part "to bring together in one volume some of the major theories of personality organization . . . " (p. xiii), very little use is made of theories or experimental devices which do not imply either a factorial approach or a trait-type interpretation. Such organismic or "specificity" studies as do appear are employed mainly as foils for a dimensional point of view.

The preliminary anticipation—and the concluding assertion—made by Eysenck is that an examination of the seemingly contradictory data in factorial explorations of personality actually demonstrates a basic agreement. The evidence for many different personality dimensions is traceable mainly to semantic disagreements, inadequacies of the factor techniques employed, and the differences in aims of respective investigators. That these differences are peripheral is made clear by reference to a hierarchical structure of personality organization—from specific responses, through habitual response and trait levels, to a type level. At this highest organizational (type) level, he believes that three orthogonally related dimensions are implicitly agreed upon—Neuroticism, Extraversion-Introversion, and Psychoticism. Eysenck does not imply that these are the only higher-order noncognitive factors. In fact, he states: "Nor can it be maintained that these are the only higher-order factors which may be discovered; nothing can be said yet about the total number of such factors required" (p. 319). Yet, curiously, in discussing second-order factors involved in social attitudes, he strongly denies the likelihood of their being more than the R (conservative-radical) and T (tough-tender minded or practical-theoretical) factors already determined (p. 244). Such

assurance seems out of keeping with the scientific spirit so readily invoked elsewhere by the author.

Eysenck's books, considered together, constitute an extraordinarily well-documented and tightly woven thesis. Although in reading them one encounters a surfeit of "factors" and "dimensions," there is no doubt that the internal reasoning is both challenging and sound. They suffer particularly from a paucity of "external" (e.g., anthropological) supportive material such as distinguished Cattell's (1) factorial approach to personality. For those who are factorially inclined, the system outlined may be completely acceptable as a pathway to personality. For others, its monorail construction may prove to be discouraging. Whatever his point of view, no psychologist can afford to ignore it.

Vernon—*Personality Tests and Assessments*

After the supercharged atmosphere engendered by the Eysenck books, this slim volume comes as a cool and refreshing breeze. It covers almost exactly the same ground as Eysenck's (1953) review of personality tests, but it does so without restricting the material to factorial studies exclusively. Although the book is only 200 pages long, it reports succinctly on attitude and interest measurement, ratings and questionnaires, physical and behavioral tests, and interview assessments. In addition, there is material on cognitive tests, and a chapter on projection (English nomenclature for the American "projective") techniques.

Vernon reports and interprets data in a calm and studied manner. Because of its brief but unhurried digestion of assessment devices and its cogent but dignified criticism, this book is recommended to the students and scholars alike who desire an unbiased orientation to the field of personality testing. It brings no solace to those who seek vindication of one or another technique, but treats all techniques as legitimately adventurous early steps in the learning-to-walk period of valid assessment.

Vernon's book is written from what he calls a "trait point of view," but not so rigidly as to make him an idol worshiper. He accurately points out the difficulty of selecting distinctive syndromes of behavior for criterion purposes, the unreliability of psychiatric and associates' ratings used for the same purpose, and the probable existence of traits which "cannot be anchored to any syndrome" (p. 11). Even more important, he recognizes the artificiality of studying personality in isolation from society.

Vernon's contentions, that " . . . an infinite number of solutions exist to any factorial problem" (p. 12) and that "no two leading psychologists agree as to which are the best dimensions" (p. 12), are in marked opposition to Eysenck's position on both points. So, too, is his belief that *Q-* and *P-*techniques are especially fruitful. The two authors do, however, find common ground in their agreement upon the subjectivity and unreliability of the "clinical" approaches. Eysenck would discard them entirely on this account, but Vernon argues for the use of objective tests and techniques to back them up or correct them.

Vernon's chapters follow a fairly consistent pattern of concisely reviewing subtechniques which fall under selected rubrics; each includes clearly labeled discussion sections when the author feels they are warranted. An example of this procedure, well worth reading, is the

discussion of paper-and-pencil personality tests in the chapter, "Self-Ratings and Personality Questionnaires" (pp. 136–140).

His eight-page concluding chapter, "Conclusions and Future Developments," is a masterpiece of fulsome brevity. It conveys neither the extent to which his summaries cut cleanly to the heart of the evaluation, nor the number of research programs implicitly designed. Its three sections deal with the situations in which personality assessments are principally required: selection, experimentation, and diagnosis or guidance. The use of personality tests in selection procedures is the most successful, Vernon feels, because theoretical problems of personality are at a minimum, and external validation criteria are readily available.

From the standpoint of diagnosis, guidance, treatment, and control, however, the reverse is true. Theoretical problems are maximal, and criteria—both internal and external—are generally lacking. For those who are familiar with Vernon as a protagonist of factorial and trait-composite research, his recognition that they are cross sectional, and not dynamic, may come as some surprise. Even more unexpected, perhaps, is his hope "that a more fruitful system will eventually replace [trait psychology]" (p. 205).

His final plea is expressed as a need for the direct recording of behavior. This, he feels, will reconcile trait psychology with field theory, with the study of social groups and processes, and with longitudinal studies of personality development.

Nowhere, perhaps, will the American psychologist be brought face to face with one of his own problems more clearly than in this book. His English colleagues have observed and have often pointed out the limitations of clinical techniques, such as interviews, Rorschachs, and ratings, when used as everyday diagnostic tools. Although the critical studies most often cited have been made in this country, they have had less influence here than abroad in actually changing practice. It appears to be difficult for at least two English psychologists of repute—Eysenck and Vernon—to understand how we can, at the same time, be scientifically self-critical and clinically self-indulgent. This trait composite has yet to be explored by the culture-oriented psychologist.

Spinley—*The Deprived and the Privileged*

Spinley's book serves to show how a good idea can go wrong. The idea itself is symptomatic of one probable trend that personality psychology can be expected to take in the future. Two hypotheses imply, roughly, the direction of this development:

1. Groups exist which can be distinguished by their material and social environments.
2. Over and above individual differences, there are "personality types" which can be associated with, and attributed to, specific cultures or subcultures.

Spinley would add to these a third hypothesis, that "there is a basic personality type; therefore the groups will possess some cultural aspects and some personality trends in common" (p. 18). Her book, a modification of her doctoral dissertation, attempts to substantiate these statements.

The populations chosen for experimental verification consisted of two English subcultures: one was a "slum group," the other a group of economically privileged graduates of English Public Schools. The author unfortunately obtained her data from the two groups in quite different

ways. Among the slum dwellers she played the role of participant observer in successive age-group clubs in a settlement house. In addition, she utilized structured interviews with informants selected from among social workers and mothers. The privileged group data consisted of 42 solicited life histories that were supplemented by interviews with nurses-in-training to establish optimal child-rearing practices. Rorschach tests were administered to samples of 30 people from each of the two subcultures.

One of the most informative chapters is that in which data from the deprived group are collated. A series of 117 items characterizing this group are arranged in maturational order, and each is coded according to its source: statements based on field observation and supported by the large majority of informants; statements based on field observation, but with the opinion of informants fairly evenly divided; statements based solely on observation; and statements based solely on data from informants. The data from the privileged groups are extracted from the life histories and hence do not necessarily overlap the statements about the deprived group. Nevertheless, the author derives a series of characteristics that purportedly typify the personalities within the respective groups. In armchair fashion she then attempts to reason out the culture-causal factors producing each. Finally, having administered and scored the Rorschachs herself, she finds by comparative analysis that her data are consistent in differentiating privileged from deprived personalities as she depicted them.

Spinley's book could well be labeled a study in unreliability. Neither the instruments nor the judgments are shown to be reliable. Her own reliability is not guaranteed, and her role introduces the dangers of halo effect and unintentional bias. It should be noted that Spinley is aware of these sources of methodological impurity, but pleads the limitations of time and resources of the lone investigator.

When objectives rather than methods, are considered, it would appear that Spinley in some respects had taken her cue from Vernon in looking for socially induced personality syndromes. She is, of course, not the first to do so, nor is her approach essentially different from the workaday methods of sociology and cultural anthropology. To the "personality psychologist," however, a change in objective might be desirable. Perhaps this could be an emphasis on socially induced *perceptual* prototypes rather than on socially induced behavioral prototypes. Better still, we might hope to trace the relationship between the objective social situation, the perception of that situation, and the behaviors which follow from it. In this way, we might crystallize the functional interplay of the many disciplines which study society and the person.

Notcutt—*Psychology of Personality*

Among the books reviewed here, this is the only one which expressly attempts to present the gamut of contemporary theories of personality, including the techniques used, the results obtained, and the applications thereof. The categorization of personality theories that Notcutt employs as a point of departure leads one to expect a thorough and uniform development of each, somewhat in the manner of Hilgard's treatment of learning theories. Instead, one finds a hodgepodge of historical reflections

interlarded with tabulations of basic motives, personality types, and defense mechanisms.

In twelve extremely short chapters Notcutt covers the history, interaction, and interpretation of seven theories, two major assessment methods, and the logic of validation. The coverage, as might be expected, is spotty and thin. The depth of treatment may be evaluated by quoting briefly:

> It is said that in some parts of the United States morticians employ minions to listen in to the short wave police radio, so that whenever a fatal accident is reported, they can hurry to the spot and sell a funeral. If psychologists had shown more enterprise of this kind, the subject might have progressed faster (p. 185).

In short, Notcutt's abortive analysis of personality theory is not recommended reading.

Nuttin—*Psychoanalysis and Personality*

In recent years there has been considerable conflict between psychoanalytic theory and certain religious beliefs. Nuttin's book has the distinction of being among the first to deal with this problem by reinterpreting and extending a psychoanalytic point of view, rather than merely opposing it.

Nuttin gives an unprejudiced account of Freudian theory in the first chapter, but gradually and gracefully weaves into it the points he is to use later in refuting its application to "normal" behavior dynamics. Thus, Freud failed to examine seriously the role of the conscious ego; Freudian theory is inadequately bolstered by experimental evidence; psychoanalytic *theory* is not inseparably connected with psychoanalytic *therapy*. Arguing that psychoanalysis is not the only psychotherapeutic

method, Nuttin points to Rogerian nondirective therapy as a desirable alternative. This particular choice supports the position that such therapy "permits the patient to construct his own personality" (p. 91 and that it is based upon a "social" rather than a directive relationship.

Nuttin is now ready to reject unconscious motivation in favor of freedom of will as a major determinant of **normal behavior. An interesting** point in his thesis is that very little human behavior is *not* unconscious, whether it be of psychic or biological origin. What is crucial, he says, is not the source of needs, but the felt need and its influence. It follows from this, he continues, that the psychology of normal people need not be based upon repression but upon the law of effect. Thus normal behavior develops according to success achieved and consciously known. Undesirable behavior disappears because "it does not continue to bring satisfaction to the whole personality" (p. 188). Nuttin completes his argument with the postulation of three levels of human activity correlated with basic needs: a psychophysiological level, a psychosocial level, and a spiritual level. The last level is coordinated particularly to man's need for self-preservation and self-development.

Nuttin has evolved a theory of normal behavior to fill the gap presumably left by those concerned with pathology alone. In so doing, however, he creates a dichotomy which is untenable in psychological theorizing. His highly speculative and metaphysical reasoning is supported only by an occasional clinical interview. His conclusion is completely predictable. Nonetheless, his attempt to cope with, rather than oppose or proscribe, "controversial"

material is a refreshing change. It seems likely that Nuttin's book will find wide acceptance in those groups which need to reconcile psychotherapy and religious faith. Despite its persuasiveness, it is not likely to gain wide acceptance among personality theorists.

Jersild—*In Search of Self*

This book by Jersild—the first of the American authors represented here—is a remarkable reiteration of Nuttin's premise regarding the behavior of normal people. Nuttin argued that people operate in the direction of constructive self-realization, and achieve it through the operation of the law of effect. Jersild, influenced by Lecky (3) and Harry Stack Sullivan (5), holds similarly that "a person accepts and incorporates that which is congenial to the self-system already established, but he seeks to reject or avoid experiences or meanings of experiences which are uncongenial" (p. 14). Unlike Nuttin, however, he believes that people may show either self-acceptance or self-rejection, the former being essential to mental health, the latter its unhealthy counterpart.

In Search of Self reports a fact-finding procedure designed to help teachers better understand their students and to assist them in achieving self-acceptance. It consists of a tabulation of items reported by students in written essays on "What I Like About Myself" and "What I Dislike About Myself." The responses of almost 3,000 individuals, ranging from fifth-graders to college seniors, are reported according to grade level, sex, and type of response. Unfortunately, statistical information is limited to frequency and percentages of individuals reporting in each category. Although the data

actually suggest little more than that young people may like or dislike themselves for almost an infinity of reasons, Jersild builds precariously on this sparse evidence in his discussion of categories of self-acceptance and rejection. He could readily have made his appropriate and instructive remarks without reference to the research data. His most valid generalization, labeled the "universal language of the self," is the recognition that "categories of self-evaluation at any one age level are also prominent at other levels" (p. 30).

The contribution of Jersild's book may not be unique from the standpoint of personality theory, but it is certainly a timely and well-developed contribution to educational theory and practice. Part Three of the book is devoted to the problem which the school and teacher must both meet—an adequate setting for the individuality of student needs, with particular reference to self-acceptance. To this point, he says:

The failures, reminders of limitations, and the rejection which children face at school are often artificial and forced. They may have the effect of humiliating the child by depreciating his worth in a manner that does no good to society and does him great harm. Much of the failure at school is contrived. Much of the depreciation children encounter there is based upon false evaluation (p. 91).

With regard to educational "devices," Jersild points out the uncertainty of their results. In group discussions, for example, anxiety may not be released, but by-passed in intellectual or logical flight. On the other hand, individuals may betray by their exaggerations, not so much their thoughts as their feelings on the subject under discussion. It becomes the job of the teacher to appreciate the individual effect and to shepherd the student's struggle to be himself. In so doing, the teacher will

himself gain opportunity for productive rather than conforming behavior. In addition, he will be better able to cope with his own problems, particularly those evoked by contact with students.

Finally, Jersild clarifies the position of the teacher as a psychologist. He is not, of course, to take on the role of the professionally trained psychologist or psychiatrist. However, "every teacher is in his own way a psychologist. Everything he does, says, or teaches has or could have a psychological impact. What he offers helps children discover their resources and limitations. He is the central figure in countless situations which can help the learner to realize and accept himself, or which may bring humiliation, shame, rejection, and self-disparagement" (p. 125).

In summary, Jersild's book has the virtue of attempting to translate personality theory into an action level. Those who operate in the rarefied atmosphere of theory for theory's sake may find it speculative and wandering at times, but will not reject it. On the other hand, the classroom teacher for whom it is surely intended and appropriate may find it too abstract and theoretical to follow. Perhaps its optimal place is in the hands of those who teach teachers, as a working guide to a sophisticated educational philosophy.

Harrower—*Appraising Personality*

This book was ostensibly written to acquaint the medical practitioner with the nature of psychological testing and its value as an adjunct to diagnosis and treatment. More obvious is its attempt to establish the respectability and validity of certain projective techniques, particularly in the hands of a trained psychological interpreter. The devices with which Harrower is concerned are the Szondi test, the Miale-Holsopple sentence completion test, human-figure drawing, Wechsler-Bellevue response "scatter," and the Rorschach test.

The method of presenting these devices is an unusual one—a not very Socratic dialogue between a physician and a clinical psychologist. The physician is no ordinary general practitioner, being much too well primed. He asks about the Szondi pictures, "What are these strange looking faces?" and later, "In what way may I be legitimately paranoid? What does it show if I like the paranoid faces?" The psychologist replies with characteristic clarity, "You would accept consciously the need to be emotionally driven beyond yourself, to become involved in or involved with other things or persons and by so doing extend your own frontiers and boundaries" (p. 126). It is unfortunate that instances of this type of dialogue occur all too frequently, the psychologist speaking with equal assurance on each such occasion. It is possible to summarize his confidence (Harrower identifies with a male psychologist in this book) on the basis of the statement " . . . by and large, psychological tests have the same margin of error as do those concerned with somatic phenomena." She is referring specifically to the 10 per cent of false positives on the Wassermann test, which was taken as a sample of somatic diagnostic techniques. It would seem that Harrower's interpretation of validation studies is at great variance with that of Vernon, Eysenck, and most American psychologists. While all might place credence in the ability of these devices to mirror some extremes of personality, their present determinable

accuracy for an unselected group would certainly be less than 90 or even 50 per cent.

Harrower does present one area with more restraint, viz., her interpretation of "scatter" on the Wechsler-Bellevue. Especially good is her review of responses to single test items, particularly with respect to the clinician's opportunity to make judgments by observing the patient in action. It seems quite likely that much more of the diagnostic success of these techniques is attributable to such observation than to the test indices and scores themselves. This point is corroborated in Harrower's (synthetic?) cases reported in testimonial fashion in the last section of her book. A number of physician referrals are seen by the psychologist, who reports back to the physician the psychological status of the patients. The first of these—under the title "Which Twin Has Epilepsy"—presents no problem if one ignores the test results but reads carefully the behavioral description. This reviewer, at least, was able to make an educated guess (a correct one), and presumes that with prodding he could also have told which twin had the Toni.

The next case deals with a "backward" girl, about whom the physician inquires, "Would you give me your opinion as to the chances of her success in a therapeutic venture?" The answer would be obvious to most who had determined her verbal IQ to be 50. With Harrower, it requires the combined virtues of Szondi and Rorschach tests to divulge a psychotic ego picture, and hence contraindicate therapy.

Psychologists of logical bent know that somewhere they must *hold* the line, theoretically speaking. This book, although certainly written from honest convictions, is apt to designate where they will *draw* the line, instead.

Blum—*Psychoanalytic Theories of Personality*

Blum's avowed goal is to supply a framework for psychoanalytic personality theory since it is at once the "most comprehensive one of its species" (p. viii), and has fared the best in practical applications. Although we may question these premises, it is apparent that the author has summarized exceedingly well the many and varied psychoanalytic points of view represented. Regrettably, the coverage is more extensive than intensive, and its structure encourages browsing curiosity rather than intellectual involvement.

Each chapter of the book represents a segment of the chronological sequence of personality formation, psychoanalytically conceived. It thus begins with "Prenatal and Birth Influences," includes the "Latency Period," and ends prematurely with "Adult Character Structure." Within each chapter, subsections are devoted to ego and superego formation, psychosexual development, relations with others, and mechanisms. The subsections are in turn subsected (an induced neologism) to note the positions taken by Freudians, neo-Freudians, and other psychoanalytic theorists. Each chapter ends with a summary and author's notes; the latter consist of reports of experimental research presumed to be related to the chapter material, suggestions for further research, and the author's critical comments.

The organizational structure of the book seems to defeat its purpose in a number of ways. The persons for whom the book was written—graduate and undergraduate students in

psychology and related disciplines—will not enjoy reading it through because of the difficulties of assimilation inherent in its discontinuity. If they attempt to trace a single concept such as "libido-regression," they will be forced to turn repeatedly to the index and to the respective sections where the term occurs. To compare one such concept with another makes the process doubly inconvenient.

For research workers, the structure is equally hampering. Blum, in his notes on each chapter, summarizes a variety of recent psychoanalytic research materials. The summaries, however, are so embedded in other material and so far removed from the theoretical content to which they refer that they appear to be superficial and unnecessary. The major accommodation for the research worker is actually the 295-item bibliography.

Patty and Johnson —*Personality and Adjustment*

Any textbook on personality must comply with unusual demands. It must not be too lofty, erudite, or theoretical—nor too elementary, vague, or contradictory. It must give the essentials of theories without vitiating them, and it must do so in readable fashion. It must, above all, enable the student to see the relatedness of theories to the people from whom ultimately they are derived and to whom they will apply.

Patty and Johnson fall far short of these utopian goals and some lesser ones. At first glance, the Table of Contents gives promise of a well-structured approach to mental hygiene, using personality theory as a vehicle. But the development of the latter to this end is certain to be meaningless or misleading to the student. An example of this comes

early in the chapter on motivation, where Lewin's topology is introduced:

This viewpoint made extensive use of *field theories*, by means of charts showing arrows (vectors) symbolizing "behavior tendencies" or "pressures." ... It stresses the fact that animal behavior seems to maintain a sense of direction in a complex situation. Its weakness lies in the difficulty, if not the impossibility, of describing behavior in sufficient detail by means of the available symbols (pp. 49 ff.).

It is a rare student who will derive from this any concept of Lewin's dynamics, or do more than reject them. It would be far better to have left it unsaid.

The frequency of misconceptions and misdirections is so great as to make recasting unbelievably difficult for the instructor and impossible for the student. The chapter on perception, a keystone in the author's development of behavior determinants, consists of exactly two topics: illusions and "set," and projective techniques. Chapter 7 contains this unverifiable datum:

It appears that mental ill-health in the expectant mother affects the development of the child in ways which may appear later as anemia, nervousness, underweight, and other equally general conditions (p. 160).

There is little use in multiplying these instances, which are primarily supplied to let the reader make his choice of text both forewarned and forearmed.

Epilogue

In these last remarks, I should like to claim the privilege of forsaking the editorial third-person and give my pen a reprieve from conventional expression.

The lot of the reviewer, like that of the policeman, is not a happy one. When he wants to say, belligerently, that the author tried to stuff a

rhinoceros down a mouse hole, he feels restrained by the knowledge that those who criticize reviewers are rougher than those who criticize books. If he attempts a bit of whimsy, he may be accused of being precious or puckish. If he's for something, he's really against someone. It is probably thus that reviewers develop tough hides that cover once tender hearts.

In looking back on these reviews, I am struck by their general negativity. I do not think that this comes from looking too long through jaundiced eyes. I am inclined to believe that for books on personality the year has indeed been fallow and lacking in conceptual breadth. But Pollyanna-like, I feel that it represents only a breathing spell, a period in which we consolidate our gains and prepare to strike out in new directions. It was for this reason that I characterized our psychological course as epicycloidal, and our position as nodal. It is fortunate that the analogy stops here—for the epicycle comes back to its starting point, while we may, with scientific outlook, hope to approximate true rather than apparent motion.

REFERENCES

1. CATTELL, R. B. *Personality*. New York: McGraw-Hill, 1950.
2. EYSENCK, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.
3. LECKY, P. *Self-consistency: a theory of personality*. New York: Island Press, 1945.
4. SMITH, M. B. Some recent texts in social psychology. *Psychol. Bull.*, 1953, 50, 150–159.
5. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington: William A White Psychiat. Found., 1947.

BUROS, OSCAR KRISEN. (Ed.) *The fourth mental measurements yearbook.* Highland Park, N. J.: Gryphon Press, 1953. Pp. xxiv +1163. $18.00.

Oscar Krisen Buros' *The Fourth Mental Measurements Yearbook* is the seventh in a series of publications intended to make available to practitioners in psychology, education, and related disciplines the rich resources of tests and measurements. The series began modestly with the little pamphlet "Educational, Psychological, and Personality Tests of 1933 and 1934" (issued in 1935), followed by a cumulated monograph "Educational, Psychological and Personality Tests of 1933, 1934 and 1935" (issued in 1936). A year later, Buros published "Educational, Psychological and Personality Tests of 1936: Including a Bibliography and Book Review Digest of Measurement Books and Monographs of 1933–36" which, however, abandoned the notion of cumulating the test entries. In 1938 was published, in book form, *The Nineteen Thirty Eight Mental Measurements Yearbook of the School of Education, Rutgers University,* followed in 1941, by *The Nineteen Forty Mental Measurements Yearbook,* and in 1949 by *The Third Mental Measurements Yearbook.*

The current volume is a beautifully printed and well-organized tome of more than 1,100 pages. In no way can it be compared with the first little pamphlet. The earliest three publications were primarily test bibliographies. The shape of things to come was first evident in the 1938 publication. For in that volume, Buros undertook the prodigious tasks

of getting cooperative reviews of tests (issued as separates in the English-speaking countries) and of compiling significant excerpts from published reviews of books and monographs related to tests and measurements. In 1940, he made this task his by tradition; in the present volume, he has dignified the tradition by an opulently produced book.

The 1953 volume is intended to be a bibliography as well as a review of tests published in the period 1948–1951. But Buros has continued the practice of adding to the bibliography previously unlisted items and, whenever possible, reviews of such tests published since the series began in 1933; the scope is considerably greater in the current volume, which lists, for the first time, restricted tests together with reviews of a few of them. Indeed, Buros, as well as the test organizations, is to be congratulated for removing some of the so-called secrecy from the instruments utilized by agencies like the College Entrance Examination Board, the Educational Testing Service, and the National League for Nursing Education. Further, the volume continues the practice of listing and giving excerpts from books and monographs related to tests and measurements for the period covered, 1948–1951.

In a sense, Buros' publications represent a compromise. In the early thirties he intended to develop a test-consumers' research organization. Failing to realize such an accomplishment, he devoted himself to the lesser, but very important, goal of supplying to all, particularly to test consumers, "frankly critical re-

views." Such a limitation, of course, meant that he had to become dependent on outside reviewers rather than upon members of his own research staff. During the years, Buros has had the cooperation of the leading specialists in test construction and test utilization. Nor has Buros left the task to them alone; he has supplied the bibliography of published research related to each listed test. Such cumulation of related material has been, and should continue to be, a valuable resource for the student and the scholar. For example, the cross-references for the Wechsler-Bellevue Intelligence Scale now goes to 312; and for Rorschach to 1,219. As for books about measurement, the review excerpts that he has edited and compiled make another basic contribution toward appreciating the growing sophistication of test constructors.

An editor who conceives his function to be the improvement of test making and hence the improvement of test using deserves well of all who make and use tests. Indeed it is fitting that the Division on Evaluation and Measurement of the American Psychological Association and the Psychometric Society jointly expressed their appreciation to Buros for his enterprise over the past 20 years. The fact that these two organizations memorialized Buros indicates the value and significance of the work. Indeed, the volume's deficiencies seem entirely attributable to Buros' limited financial resources rather than to any administrative or intellectual constraints. The volume, in so far as it is not cumulative, makes it necessary for the interested user to manipulate several volumes, for test reviews are to be found in the four volumes for 1938, 1940, 1949, and 1953. Perhaps the time has now

come when test listing and test reviewing can be done best through the medium of an official magazine sponsored by the American Psychological Association. Such a publication would give greater currency to the newer tests faster. With the expanding test field, it seems too long to have to wait four or five years for Buros' next volume. Indeed, if subscribers are willing to pay $18 every four or five years, there may be even more who can afford $4 or $5 every year for more current material.

Buros, in his attempt to be fair to every test maker, has "leaned over backwards." At one time, he forwarded the critical reviews to the test authors and/or publishers to make certain that the critiques were related to the facts. This apparently was so formidable a task that he abandoned it in favor of trying to achieve at least three reviews of each test by different analysts. Fair as that may be, it seems not nearly as valuable an objective as getting at least *one* review of every listed test. In a random sample of 20 entries, four were not tests at all (a manual on aphasia, a book on Rorschach, a manual for driver selection, a series of forms for employee selection); seven were of new tests not previously listed, but which remained *without* review; three were for new tests not previously listed with a review; four were for reviews of a revised or supplemented test; and two were reviews for tests which had been previously listed and reviewed. Perhaps editorial energies could have been expended to get all the new entries reviewed at least once.

It is not difficult, however, to recognize that the reviewers themselves frequently add their personal prejudices to their test appraisals. Sometimes these biases are minor ones, such as criticizing the directions,

sometimes they are major ones, such as questioning the use of primary batteries of standardized achievement tests. For one reasonably familiar with the proclivities of the individual reviewers, it is quite a confirming reaction to be able to designate the reviewer just from the review. In a field as diverse as testing, various specialists have developed special attitudes: e.g., criticism of objective scoring as restricting the scope of testing, or the impossibility of getting truly representative national norms, or the overclaims on diagnostic profiles. As is inevitable, the reviews are uneven because of the very specialization of the reviewers. In one sense, the user of the volume should have available specifications of "all the players."

Nevertheless, most of the advice is significant and cogent. A few quotes give the flavor of the volume: the test has "little to recommend it . . . in preference to many similar commercial tests on the market," or "the items in the test also have a wide range of difficulty, the earlier items being very easy; but the last, even with the answer key available, gave the reviewer considerable difficulty," or "among the major ones are: insufficient data about *what* it predicts, temptation to do elaborate pattern analyses on subjects' scores, some urban-rural inequity in the standardization sample . . . " or " . . . strong relation to the . . . [somebody else's] reading tests in form, vocabulary, and general testing procedure; and the first grade tests could well be additional forms of the [somebody else's] *Tests*," or "study of the items reveals considerable dependence upon recall of factual information. . . . " Future editions of the volume will profit from attempts to classify the *relevant* information by validity, reliability,

sampling for normative data, character of norms, relevance for specified objectives of appraisal or of instruction.

It is the lack of a comprehensive plan for coverage that makes for the primary difficulty with the reviews: they tend to be so prolix and discursive as to omit much information that a consumer should have for his decisions. Certainly, the bulk of the reviews fails significantly to meet Buros' requisite: "Reviews should be written primarily for the rank and file of test users." Too often, the reviews seem to be prepared for the reviewer's peers.

The current reviews, however, occasionally give evidence that earlier reviews have led the authors to make improvements in subsequent revisions of their tests; unfortunately, in too many instances, the reviews give contrary evidence that the test authors were immune to such analyses. Over the 20 years, however, Buros has raised the standards for the technical aspects of test making. It is hoped that a magazine may continue these important reviews to make them more up to date and available at a price that will allow the rank and file to use them more freely. I know of no one as capable as Buros to be its editor.

IRVING LORGE
Institute of Psychological Research
Teachers College, Columbia
University

MOWRER, O. H., and 21 Contributors. *Psychotherapy, theory and research.* New York: Ronald Press, 1952. Pp. xviii + 700. $10.00.

This is an important book, but, by virtue of the range of content and number of authors represented, it is a difficult one to review. Part I, entitled "Theory, Concepts and Appli-

cations," consists of eight chapters, two by Mowrer and six by other contributors. Part II is entitled "Research Methods and Results" and consists of 11 chapters. Of these, Mowrer is the author of two and coauthor of three more; the remaining six chapters represent the work of other contributors. Mowrer also provides a brief Preface and Introduction; there is, however, no attempt to summarize or integrate the many diverse contributions, which vary widely in both content and quality. All but two of the 19 chapters were written expressly for the present volume. Finally, the book is big: the 700 pages average about 500 words per page and include 21 tables, 120 figures, 211 footnotes, and a bibliography of over 300 items.

In his Preface, Mowrer states that "mental disorder and the challenge of its more effective understanding, treatment and prevention constitute one of the major preoccupations of our contemporary civilization." He observes, however, that only since World War II have psychologists concerned themselves centrally with psychopathology and psychotherapy. This volume, written exclusively by psychologists who for the most part are experienced therapists as well as active investigators, is presented as a partial report of progress regarding developments on this research frontier. Because of this general limitation to authors engaged in actual research on psychotherapy, the volume is in no sense representative of either the theory or practice of present-day psychotherapy. It does, in the 500 pages of Part II, include a good sampling of the research methods used and the preliminary results obtained by psychologists doing research in psychotherapy.

Part I begins with a delightful chapter by Rollo May tracing the historical development of social attitudes toward neurotic manifestations and the modes of treatment provided at different historical periods. This is followed by an admittedly speculative chapter by Carl Rogers in which he emphasizes the importance of examining the process of psychotherapy entirely apart from its utility, and after conceding "that the process, directions, and end points of therapy may differ in different therapeutic orientations," attempts to delineate the major common aspects of the experiences of clients undergoing nondirective therapy. In the third chapter, Mowrer undertakes to develop briefly a systematic account of both "neurosis and psychotherapy as interpersonal processes." In this paper, Mowrer continues his argument with more orthodox psychoanalytic theorists on the question of the nature of repression and its implications for therapeutic strategy in handling transference.

Chapter 4 by William Perry and Stanley Estes represents an attempt to resolve the directive-nondirective controversy by conceptualizing the counseling process as one requiring "The Collaboration of Client and Counselor"; the counselor plays two distinctly different roles in two phases of the counseling process. In the first, he should be truly nondirective, but after the client accepts responsibility for his own active participation in solving his problems, the counselor may appropriately change his role to that of an expert collaborator in the actual process of problem solving. Preliminary research findings are presented to support the validity of this orientation.

Chapters 5 and 6 are both analyses of psychotherapy as a learning process. In the former E. J. Shoben

suggests that in therapy the patient learns a new "comfort reaction" to replace his anxiety reaction and that these new reactions generalize sufficiently to extraclinical situations to permit the patient to develop new and more appropriate instrumental social acts. In the latter, Mowrer defends his modifications of Freudian theory on the basis of his two-factor theory of learning.

Chapter 7 is a somewhat modified version of Milton Wexler's previously published interesting account of his two years of therapy with a severely disturbed schizophrenic patient and a statement of the modifications in psychoanalytic treatment which he believes necessary for successful therapy with such patients. In general, his suggestions for revision of theory and practice seem congruent with those proposed by Mowrer for the treatment of neurotics, which perhaps accounts for the inclusion of this paper on schizophrenia in a book otherwise dealing exclusively with neurotic rather than psychotic behavior.

Chapter 8 is a discussion of the implications of psychotherapy for public health, by Joseph Bobbitt and John Clausen, with especial emphasis on the parallels between psychotherapy, group activities, and other educational methods which show some promise of being effective both therapeutically and prophylactically in the mental health field.

Looking next at Part II, we find first a brief but useful summary by Jules Seeman and Nathaniel Raskin of the research methods and results growing out of "client-centered research on therapy." Chapter 10 is a reprint of Dollard and Mowrer's 1947 paper describing the Discomfort-Relief quotient as a method of measuring tension in written documents;

this serves as an introduction to Chapter 11 (Mowrer, J. McV. Hunt, and L. S. Kogan) summarizing further studies utilizing the DRQ; it appears to be significantly related to therapists' estimates of therapeutic success and several objective measures of personal change during therapy. In Chapter 12, Fred Fiedler presents an integrated summary of his studies on the attitudes and feelings of therapists of different theoretical persuasions and the role of these feelings in the therapeutic relationship.

The next four chapters deal with the application of relatively new statistical procedures to clinical research. Chapter 14, written by Mowrer, is entitled "Q-Technique—Description, History and Critique." In this reviewer's opinion, this scholarly 60-page methodological monograph is the most valuable contribution of the entire volume. In it, Mowrer begins by illustrating recent applications of the technique in research on psychotherapy, traces the history of the technique and its relation to conventional factor analysis and other alternative research designs, and makes a clear-cut distinction between Q technique as a method of data collection (which he proposes to call S technique) and as a method of data analysis. Although agreeing that these "statistics of the individual" hold much promise for research on psychotherapy, he also reminds us of their limitations and pitfalls, unless used in the context of the alternate analytic procedures suggested by Cattell's "covariation chart" and by factorial design, in which case there seems to be real hope of integrating "experimental" and "clinical" types of research.

In Chapter 14, Lee Cronbach contributes a briefer and more general discussion of "Correlations between

Persons as a Research Tool" and its relations to other methods of analyzing complex data. Chapter 15, by Lester Luborsky, describes the "P" techniques (correlation of repetitive measures of the same person) and demonstrates its use with a patient in the course of psychotherapy. In Chapter 16, L. L. McQuitty describes still additional methods of evaluating "Personality Integration" and suggests their potential utility in clinical research.

Chapter 17 is another 80-page "monograph" by Mowrer. In this he first develops the thesis that neurosis may usefully be regarded as a "disturbance in communication" and that psychotherapy is essentially a matter of assisting the patient in re-establishing good channels of communication between different aspects of himself and with others. He then reviews an impressive series of empirical studies relative to language, psychopathology, and psychotherapy and suggests the use of several procedures (e.g., Osgood's "semantic differential") as promising methods for research on therapeutic change.

Chapter 18 is an even larger chapter —nearly 100 pages—entitled "Tension Changes during Therapy with Special Reference to Resistance," by Mowrer and three of his students (B. H. Light, Zella Luria, and Marjorie Zeleny). In the theoretical part of the chapter, the authors begin by reviewing the Freudian formulations regarding resistance and attempt a resolution of certain alleged "paradoxes." They next review and evaluate Rogers' views on resistance and conclude that the phenomenon is of sufficient importance both theoretically and practically to justify the development of methods permitting its objective measurement during the course of therapy. The

major portion of the chapter is devoted to reporting the development of a simple and practical technique for measuring "palmar sweating" and enough preliminary data to indicate its promise as an index of tension and hence, perhaps, of resistance. The final brief chapter, written by E. I. O'Kelly, is devoted to a survey of the possibilities and limitations of other physiological measures as indices of change in therapy.

This then is the content of the book. It reflects a tremendous amount and variety of activity on the part of a great many psychologists. And, although the volume is heavily weighted with the thinking and research methods of Mowrer and his colleagues, a reasonable effort was made to include at least samples of research by persons with other orientations. This together with the relatively undeveloped and unorganized state of knowledge in the field resulted in a volume perhaps better characterized as a "Contributions to . . ." or "Readings in . . . " than a "textbook" or "treatise" as its title tends to imply. Certain of the chapters might, in this reviewer's judgment, have been better published as separates since they bear but little relation to the major theme of the volume. An integrative summary chapter would have been helpful, but perhaps unnecessary, to emphasize the fact that we are not yet ready for such an accounting.

The major criticisms that may be made of this volume are those which are inherent in the present undeveloped state of an art which may sometime become an applied science. One finishes the book with such questions as: "What is psychotherapy anyhow? "Is it effective—and if so under what conditions and with what kinds of clients or patients?" Even when re-

search results are presented, the reader has difficulty in evaluating them because our present state of knowledge does not permit an adequate description of either the nature or intensity of the presenting problem. Furthermore, in view of the extremely wide variety of activities subsumed under the rubric "psychotherapy," the reader is at a loss (unless he can read an entire series of transcripts) to know what really occurred in the treatment hours. Finally, as this book so well shows, the problems of developing criteria are most complex. For the most part, the authors of these contributions are fully cognizant of these major problems but confident that they are amenable to research.

Whether or not the reader agrees with Mowrer's theoretical orientation, he will find this a challenging and useful series of contributions to an increasingly important topic in psychology. It is a book which no one planning any kind of research on any aspect of psychotherapy dares to overlook. I predict that but few persons will read all the chapters in sequence, but that equally few will fail to find several of the chapters extremely valuable.

E. LOWELL KELLY
University of Michigan

LACEY, OLIVER L. *Statistical methods in experimentation.* New York: Macmillan, 1953. Pp. xi + 249. $4.50.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953. Pp. xix + 393. $6.50.

WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953. Pp. xi + 510. $6.25.

Depending on one's point of view,

it may be either distressing or satisfying to note the increasing extent to which training in statistical methods has invaded the curricula of psychologists in recent years. Little more than a decade ago the typical Ph.D. in psychology was generally considered to have a respectable statistical repertoire if in addition to histograms and polygons he had acquired an assortment of measures of central tendency and variability, manipulations of the normal curve, linear correlation and regression, and the critical ratio. Sophisticated indeed were students who had mastered the intricacies of the *t* test, chi square, and partial-multiple correlations. With due respect to Truman Kelley, the beginning of the modern era, characterized by the continuous translation of newer statistical logic and procedures for the consumption of mathematically naive but eager psychologists, can probably be dated to 1940, when Lindquist's *Statistical Analysis in Educational Research* and Peters and Van Voorhis' *Statistical Procedures and Their Mathematical Bases* made their appearance.

Since 1940, and especially in the last five years, a number of "advanced" statistical texts designed for psychologists have been published, and there has been a remarkable increase in the number of statistical articles and reports of complexly designed research studies. Also, the number of statistics courses offered in many graduate schools has shown a concomitant rise. In some schools emphasis on statistics has become so strong that students are urged to enroll in courses in mathematical statistics and spend summers at statistical institutes.

Thus, today the research psychologist is expected to possess a competence in statistical design, analysis,

and inference which can be produced only by devoting a substantial portion of the graduate program to courses in statistics or by motivating students to do a considerable amount of self-study. Perhaps unfortunately, the increased emphasis on statistics seems in part to have been accompanied by a decreased emphasis on psychophysics and measurement theory. If psychologists should also be competent applied statisticians—and this is debatable in view of the trend toward teamwork in research—it would seem that a partial solution to the problem lies in selling the idea that students should acquire more relevant mathematics at the undergraduate level rather than simply increasing the statistical load on graduate students.

Lacey's book represents an attempt to bring together the basic concepts and techniques of modern statistical inference in a form suitable for the first course in statistics. The main trouble with this excellent little book is that it is so completely introductory. Beyond a compact discussion of the rationale of experimental design and inference, the specific topics touched upon are the binomial expansion, the normal curve, the t test, chi square, simple correlation and regression, and the elements of fiducial limits. It is hard to judge where this book might fit into the program for the training of psychologists. The material appears to be too scanty and pitched too low for a full-fledged graduate course. However, at the undergraduate level one might venture to replace the usual semester in reductive and descriptive statistics with a course focusing immediately on statistical inference.

Perhaps the main utility of Lacey's book will be to serve as a reference text for courses in experimental psy-

chology. In view of the current status of psychological research design, one can seriously question Lacey's claim that "when you complete the text, you should be able to read and understand over 80 per cent of the experimental articles in your field, and should also be capable of designing and executing many of the experiments which you yourself may wish to try" (p. viii).

Lindquist's book furnishes an amazingly detailed account of both the theory—reduced to algebra—and the principal applications of analysis of variance and covariance in educational-psychological research. Few psychologists will complain of a paucity of material in this book, which attempts to do for psychological experimentation what Cochran and Cox's *Experimental Designs* has done for biology and agriculture. Closely written but at the same time deliberately repetitive in the interest of good pedagogy, loaded with algebraic proofs and discussion of assumptions and pitfalls, illustrated with an abundance of actual experimental situations in both example and exercise, this book should make a major contribution to the quality of psychological research. There is easily sufficient material for a year's course in experimental design and analysis. Equally important, this will be a valued handbook for anyone engaged in psychological experimentation.

Although the entire book is unique in view of its exhaustive treatment of a multiplicity of useful designs, especially noteworthy are the chapters devoted to groups-within-treatments designs, controlling individual differences in factorial experiments through the use of "mixed" designs, and tests concerned with trend. The material on interpretation of interactions and the selection of valid error terms is

developed with outstanding skill.

This comprehensive and scholarly book would have been even better if a more adequate job had been done in referring to the relevant literature. The book is so self-contained that it tends to become dogmatic about a number of controversial issues. No book of this type should fail to convey the important concept that it is, after all, only a selective progress report on the never-ending search for mathematical models that approximate the relationships of observed data. This reader would gladly have sacrificed a number of the proofs in favor of a discussion of the problems encountered in dealing with non-orthogonal data. Although the section on Type II errors was admirably presented, it was a surprise to read the statement saying that the basis for testing Type II errors has been worked out only for simple randomized experiments involving two treatments (p. 71).

It was real joy to note the appearance of the long-awaited book by Walker and Lev. This book is a major contribution to the statistical training of psychologists. Just as Lindquist's book is unique in its treatment of analysis of variance, Walker and Lev have done a unique job in surveying the total field of statistical techniques useful to psychologists. In contrast to Lindquist's decision, the authors in this case make no attempt to supply "proofs" for procedures, but content themselves with ample discussion of the logic of underlying assumptions. The bibliographies at the end of each chapter are excellent, and the text and exercises provide strong motivation for follow-up of the literature cited.

Here again is sufficient material for at least a full year's graduate course beyond the introductory level. The combination of Walker and Lev with Lindquist might conceivably be compressed into a minimum of three semesters. Interestingly, Walker and Lev provide considerable material in the area of analysis of variance and covariance which is not presented by Lindquist.

Mere citation of the headings of the 18 chapters in Walker and Lev's book would, with the exception of the final chapter on "Non-Parametric Methods" written by Lincoln Moses, give little hint of its comprehensive coverage. Traditional material is presented in the context of the most modern of statistical logic, but in addition every chapter contains the bonus of an excellent selection of recently developed techniques for dealing with common problems that arise in the analysis and interpretation of experimental results. An extremely useful set of 24 tables and a glossary of symbols are appended.

In order to fulfill his function as critic, the reviewer feels obliged to make at least one criticism of Walker and Lev's book. In the section dealing with the comparison of two proportions based on the same individuals (p. 102), it would have been most appropriate if the reference had included McNemar's 1947 article as well as the article by Cochran published in 1950.

It cannot be stressed too strongly that the texts by Lindquist and by Walker and Lev are by far the best ones available in their respective areas of coverage. Both are based on the assumption that a statistical text for graduate students should not be a mere cookbook devised for spoonfeeding. Obviously these authors believe that research workers in psychology and education should be competent applied statisticians. The

reviewer concurs and hopes that teachers of statistics will have the courage to adopt these two books widely so that potential research will benefit accordingly.

LEONARD S. KOGAN
Institute of Welfare Research,
Community Service Society of
New York

LEHMAN, HARVEY C. *Age and achievement.* (Published for the American Philosophical Society.) Princeton: Princeton Univer. Press, 1953. Pp. xi + 359. $7.50.

Lehman's topic is one of urgent concern to psychology and to society in general. Since age is an inescapable attribute of every person, it has been an ever-present vector in determining human behavior. In view of the fact that human longevity has greatly increased in the recent past and appears likely to increase further in the near future, better knowledge of the role of age has become progressively more critical.

Age and Achievement assembles in convenient form data which Lehman and his collaborators have been collating for more than two decades. Nearly all of Lehman's previous contributions apparently have been incorporated here, although the volume makes no explicit statement on this point and provides no complete bibliography of Lehman's publications. The book consists of earlier research reports partially rewritten for this presentation and supplemented by some new data.

The majority of Lehman's researches have been directed to the production of creative works, such as paintings, poems, musical compositions, inventions, and scientific discoveries, as a function of age. His general finding is that while masterpieces and "best works" may be produced at any time during the ma-

ture years, in many fields the peak of productivity is reached in the thirties and forties, if not before 30. The number of contributions per decade per surviving contributor is greatly reduced in the later years. Thus, the period between ages 50 and 70 has been responsible for a very small part of the world's best work, except in certain fields such as astronomy and history. During the fifties the production of "best works" per living creative worker in many specialties is only 20 per cent or less of what it was in the thirties, and later productivity is even less. If we extrapolate from Lehman's data, it appears that the imminent increase in the average life span from age 70 to age 80 or beyond will result in an inconsequential increase in creative contributions of first-rate importance.

These findings, if taken at their face value, are of great importance, and careful attention should be given to their interpretation. Most of the remainder of this review therefore will be devoted to an examination of this question. No lack of confidence in the accuracy of Lehman's data is implied. His data seem to us to be impressive in their objectivity, their quantity, and their self-consistency. We are concerned solely with interpretation. Lehman himself is most cautious in this regard; it is hoped that others will be equally restrained in regard to their generalizations.

First of all, it is necessary to keep in mind that nearly all of Lehman's studies of creative contributions are concerned not with the aggregate of creative products but only with "best works" and "masterpieces." This fact introduces into Lehman's studies inherent limitations which should be recognized. By definition no man can produce more than one "best work." If this product appears early in the

se span any number of later contributions of nearly equal merit will not affect analyses such as Lehman's regardless of longevity. The comparative frequency of masterpieces at different ages therefore probably provides a very inaccurate index of the relative value of work done in different decades. Lehman's studies may show that man's "best" work is frequently done fairly early in life, but the relative value of other work is unassessed.

Lehman's finding that masterpieces in a given field are often produced in the thirties may properly be likened to a report by an anatomist or physical anthropologist that the maximum height of man is most often reached during this decade. Such a report, while interesting, would give us no measure of the degree of decrease in height which occurs in the later decades. It is true that Lehman's data show that older men produce fewer works reported to be of high value than do men in their thirties. But whether the value of the septuagenarian as a creative worker relative to the 30-year-old is 99 or 50 or 10 per cent cannot be determined by studying masterpieces.

Although Lehman's primary concern is with the relative frequency of the notable contributions at different ages, in a few instances he presents data on the *total* contributions, regardless of value, made in different decades of life. In contrast to the "best" contributions, the total numbers of creative works show little decline with age. This difference between the age trend for total productivity and the age trend for highest quality productivity is clearly one that deserves additional analysis in the future.

In selecting "best works" for study, Lehman has carefully avoided the introduction of any bias of his own into his data. Lists of the outstanding achievements in any field have, in each case, been made by others, not by Lehman. But this procedure, while commendable, does not avoid biases which may have been present in the evaluations of anthologists and historians, whose decisions are utilized by Lehman. Thus, if anthologists tend to choose the earlier of two works of almost equal merit produced by the same creative worker, this tendency would obviously produce an overestimation of the productivity of masterpieces in the early decades of life. We have no evidence concerning the presence or absence of a bias favoring selection of early works on the part of editors, compilers, biographers, etc.

Some of the anthologies and histories used by Lehman as sources of data were published shortly after the death of many of the men whose works are evaluated therein. There exists the possibility that historians and editors may hesitate to include the later works of a recently deceased man because of the difficulty of judging recent contributions. It is our impression that hesitancy in making judgments concerning the recent past is common among historians. If this is true, it would lead to low scores in later maturity for the men who died only a few decades prior to the publication of a source book or history.

In some analyses Lehman treats separately men of former centuries and men who lived recently. Men of former centuries show less decrement with age than do recent men. This fact is susceptible to the interpretation that there is a bias against the citation of a man's later contributions, on the part of authors writing soon after his death.

Further influences to be considered in interpreting Lehman's data are the possible changes in standards of com-

parison that may take place in the course of a life span. Each successive quarter of a century within the past 200 years has witnessed a great increase in the ranks of creative workers and in the number of entries in lists of notable contributions. This means that a man in recent centuries has had many more competitors for honors in his field at age 75 than he had at age 25. For example, in our own field of psychology, a man publishing today who wishes his work to be noted is potentially competing with a large proportion of more than 11,000 other psychologists. Had he been writing 50 years ago, he would have been competing with a total population of scarcely more than 100. A comparable change has occurred in most other fields. As suggested above, this great increase in number of workers introduces the possibility that the standards for "best works" in various fields have been progressively raised. Lehman's data may reflect not age changes but, instead, changes in the requirements for recognition that have occurred in the life span. Analysis of data for an era not characterized by an increase in the number of creative workers might give quite different "age" trends. Evidence on this point is lacking and, indeed, may be impossible to obtain. But the absence of such evidence should make us wary of interpreting Lehman's curves as representations of changes in the individual during the age continuum.

We have noted previously that Lehman presents data showing that scientists, inventors, authors, and other creative men born a century or more ago yield age curves, in regard to "important" works, somewhat different from the age curves produced by men born more recently. In general, the men of earlier centuries maintained a better record in later

maturity than have men born more recently. Several interpretations are possible. One was suggested earlier in this review. Another is that standards of comparison in former times changed less drastically during the course of a man's lifetime than they do today.

Lehman has done a major service to psychology by assembling a multitude of facts relative to age and achievement. There are limitations inherent in his data, and we have indicated factors other than age which may be responsible in part, or in toto, for his "age" curves. Nevertheless, he has supplied the data which are fundamental to further studies. As is usual with those who pioneer in new fields, Lehman has posed more problems than he has solved. Lehman's own discussions, as well as the fine foreword by Terman, provide an excellent précis of the further research problems in regard to age and achievement which await solution.

WAYNE DENNIS

Brooklyn College

GRAY, J. STANLEY. *Psychology in industry.* New York: McGraw-Hill, 1952. Pp. vii+401. $5.00.

There has been a tendency in the past few years to rush the publication of textbooks in industrial psychology. This is true of texts in readings in industrial psychology as well as of those dealing with human and industrial relations. As a result, the selection of topics and the thoroughness with which they have been written are somewhat haphazard. Gray's text is no exception.

The book is written in a strictly logical order. In Chapter 1, the author introduces the reader to some basic concepts in industrial psychology and then proceeds to consider the job and its basic requirements. He

considers next the usual areas of study in industrial psychology, i.e., training, motivation, morale, accidents, etc. A chapter on the age of workers is also included. Orthodox presentations of these topics are made. Little attention is given to newer applications of social psychology to industrial personnel problems.

Chapters on methods of work, wages and job evaluation, nutrition and rest, and lighting and ventilation are also included. It appears that these areas should be left to the industrial engineer and more time spent on the basic psychological problems of industry. A text is needed which will start out by stating the basic objectives and functions of the industrial psychologist. Such an approach would help the beginning student and the professor who has to teach the course. Moreover, emphasis needs to be placed on methods and procedures through which industry might benefit in worker efficiency.

Gray has written a textbook which provides the basic fundamentals and concepts of the field of industrial psychology. In this respect the book is a good compendium of facts. The text, however, falls short in the presentation of the real problems of the industrial psychologist.

D. J. MOFFIE
North Carolina State College

DANIEL, ROBERT S., & LOUTTIT, C. M.
Professional problems in psychology.
New York: Prentice-Hall, 1953.
Pp. xi+416. $5.50.

"Professional" is used broadly to include the problems of the teacher, research worker, and scholar in psychology as well as the problems of the psychologist who earns his living in clinical, industrial, or some other form of psychological practice. For the members of that broad audience Daniel and Louttit have collected a surprisingly varied amount of information, much of it of course from other sources, but in good part original.

Do you want to know the names and publication dates of bibliographical sources to psychological or general literature, the address of a publisher, where you can get a fractional horsepower motor, an electronic regulator, or a test? Are you interested in a definitive list of the 331 psychological journals that have been published at one time or another, in an authoritative guide for abbreviating journal titles, or in information on the preparation of various kinds of manuscripts?

Do you want to know who attended the first APA meeting, how many members the Association had in 1943, where and when the first International Congress of Psychology was held, or the names of the four psychological journals that were published before 1800? Or do you want a list of state psychological associations, information about ABEPP examinations, or a discussion of public relations in psychology?

This sampling gives an idea of the wide scope of information contained in the four sections and four appendices which make up the book. The first section is introductory. Then follow series of chapters on psychological literature, reporting psychological research, and problems of the professional psychologist. The appendices are an annotated list of reference books of value in psychology; a bibliography of journals in psychology; a list of sources for books, tests, apparatus, equipment, and supplies; and a glossary of abbreviations useful to the psychologist.

The inclusion of sections on psychological literature and the reporting of psychological research may come as a surprise to some readers who interpret the title as dealing only with licensing, ethics, interprofes-

sional relations, and other such "professional" topics. But the inclusion is a valuable one, and the authors are well qualified in these areas. Louttit, in particular, has long been one of psychology's experts on psychological literature. Both of these sections are necessarily condensed, for they cover a lot of ground. But both are authoritative. The final section, on problems of the professional psychologist, is also brief; it too covers a lot of ground, but the nature of the material does not permit an authoritative treatment. Psychologists have not yet reached full agreement on many of the problems discussed in this section, and the very nature of the problems makes it impossible for the authors to be as definitive as they could be in the earlier sections. The major problems are introduced; some of the principal points of view are outlined; and references to fuller discussions are plentiful.

For the graduate student the book consists of a thoughtful compendium of advice about collecting a working bibliography of source materials, about the publication policies and the kinds of articles accepted by each of a number of journals, about the preparation of a manuscript and how to see it through the press, about ethics, interprofessional relations, psychological organizations, the teaching of psychology, and other problems that he will encounter as a professional psychologist. For the mature psychologist the book's chief usefulness will be as a reference volume. He already knows some of the things included, so can browse through it to get acquainted with its coverage. Then it belongs on the handiest shelf, for it will be a frequently consulted reference source.

DAEL WOLFLE
Commission on Human Resources and Advanced Training

CABOT, HUGH, & KAHL, JOSEPH A
Human relations. Cambridge, Mass.: Harvard Univer. Press, 1953. Vol. I. *Concepts.* Pp. xxxi+333. $4.75. Vol. II. *Cases.* Pp. viii +273. $4.25.

These two volumes by Cabot and Kahl constitute the fruits of their efforts to study and teach human relations at Harvard University during the last decade. Simply stated, their approach involves two fundamental views: (*a*) that effective face-to-face human relationships can only emerge from a knowledge derived from the application of fruitful social science concepts to concrete social situations. Ostensibly, a full understanding of these relationships requires not only meaningful conceptual tools but also the skills necessary for their utilization in specific social settings. And (*b*) that the communication or learning of the concepts and the skills necessary for their application can best be achieved by the medium of group discussions of case material representing concrete social situations.

Volume I is devoted almost exclusively to the interpretation and elaboration of selected social concepts, whereas Volume II consists of 33 cases of social interaction in face-to-face situations. The presentation in Volume I takes the form of a series of discussions by the authors in which the theoretical views of various prominent social theorists and practitioners are summarized and, to some extent, integrated. At the end of each discussion, i.e., chapter, many of the writings from which these views were taken are presented as a series of readings. In some cases complete articles are reproduced and in others pertinent selections from books are given. Approximately two-thirds of the first volume consists of these readings. Because the thinking of so many

theorists is involved, and inasmuch as many of these views have been extensively reported on and evaluated in the literature, critical evaluation of specific theoretical approaches will not be attempted. As a source book of human relation concepts intended for the student and nonacademic reader, the presentation of these concepts in terms of clarity and level of discussion becomes the critical question.

It is clearly evident that an attempt to understand human interaction in specific face-to-face situations implies an idiographic or "clinical" approach. The first two chapters of Volume I are devoted to a discussion of the nature and difficulties involved in such an approach. Here the authors stress the importance of theoretical generalizations in understanding the unique event as well as a consideration of all the factors in the situation and of the relationships among these factors. They caution against distortions in social observations introduced by "habits of thought and language." In effect, the discussion summarizes the views of A. J. Lowell, L. J. Henderson, C. H. Cooley, S. I. Hayakawa, and others.

Chapters III to VII deal with concepts relevant to understanding the behavior of the individual in relation to others. Most of the theoretical conceptions considered have been cited extensively in psychological and sociological literature and thus need only be cited in passing. Chapter III, entitled "Cultural Values and Social Roles," draws heavily upon the views of Durkheim ("Anomie"), Florence Kluckhohn ("Dominant and Variant Cultural Value Orientations"), Linton ("Status and Role"), Benedict ("Continuities and Discontinuities"), Parsons ("Age and Sex in Social Structure"), and Warner and Lunt ("Social Class and Social Struc-

ture"). A discussion of the conflict between the need to belong and the individual's desire for independence in Chapter IV briefly summarizes the thinking of Erich Fromm. A number of the writings of Piaget constitute the framework for a discussion of "Our Developing Social Interactions" in Chapter V. Chapter VI considers the nature of conscious and unconscious thought processes (Freud, Mumford, etc.); and finally, Chapter VII is devoted to a discussion of the effects of anxiety and fear on social interaction (Mayo, Horney, etc.).

The remaining six chapters are focused more or less on the problems of group structure and function. Chapters VIII and IX deal with "Group Membership" and "Group Processes" respectively (Cantril, W. F. Whyte, Davis, Homans, etc.). Chapter X deals with the problem of differences in group values reflected in the conflicts between racial and religious groups in the United States. The writings of Rose, Myrdal, Lewin, Frenkel-Brunswik, and others form the basis of this discussion. The problem of modifying values and feelings is discussed in Chapter XI. Here the importance of a nondirective approach in a face-to-face situation is stressed as the primary condition for bringing about change. As one might expect, the writings of Carl Rogers form the basis of much of this treatment. The final two chapters, XII and XIII, deal with the problems of "Executive Leadership" and "Social Control and Equilibrium."

It is readily patent that the concepts discussed in Volume I—as the authors themselves point out in the Introduction—are "interrelated concepts from several . . . disciplines." In various places the interrelatedness of some of the concepts is stressed, but, as the authors indicate, no attempt is made to put forward a

"tightly reasoned system of social theory." The order in which the concepts are presented rests on the assumption that effective action in a social situation must be preceded by skillful diagnosis. Thus, the discussions in the earlier chapters orient the individual toward the problem of social observation, while the later ones are more "action" oriented. Cabot and Kahl are by no means dogmatic about the order of concept presentation; they suggest that the order can and should be varied to meet the needs of particular groups.

In the main, the selection of concepts and the order in which they are presented reflect very careful thought. In reading through Volume I, one cannot but be impressed by the authors' cogency in organizing social data. Yet, one rather serious criticism of the volume should be made. We have no quarrel with the particular social concepts selected, but rather with the conspicuous absence of any systematic discussion of the biological contributions to personality development. True, the purpose of the writers was to provide the reader with a social theoretical framework for understanding social interactions in concrete situations. But they have ignored their responsibility of not leaving the student and nonacademic reader with the distorted view that human reactions are purely a function of sociocultural factors. Some preliminary discussion of the contributions of essential structure and equipment in producing differences in intelligence, drive intensity, and the like would have gone a long way toward reducing the possibilities of such distortion. While it is critical for the reader to understand the importance of learning in the development of social interaction, it is no less important for him to understand the limitations imposed on such learning

in an individual with an IQ of 70.

Perhaps the most disturbing aspect of Volume I is the level of presentation. Both volumes are clearly intended for undergraduates in integrated social science courses and nonacademic groups of various kinds. Somewhere along the line the authors lost sight of some, if not all, of their intended readers—insofar as Volume I is concerned. In general, their discussions tend to be abstruse and, in the reviewer's opinion, will tax the comprehension of the average undergraduate and certainly individuals in many nonacademic groups. In part this is because, although the writings of many of the theorists demanded clarification and a simpler presentation, Cabot and Kahl simply summarized them in much the same language. While various Steig cartoons presented throughout Volume I serve to suggest a concern with the communication of ideas, the actual discussions stand in sharp contrast to these ingenious drawings. A greater concern for the general reader by avoiding esoteric discussions would have gone a long way toward improving the value of the volume.

On the other hand, the selection and presentation of the case material in Volume II are outstanding. The cases depict actual social interactions encompassing a wide variety of situations. Inasmuch as they are to serve as the "raw data" that are to be organized in terms of the concepts presented in Volume I, they are straightforward factual accounts. They are presented in detail, with both background material and verbatim conversations and, more significantly, without any attempt to illustrate particular theoretical points of view. Because their approach to increasing the individual's understanding of human relations rests on his learning through group participa-

tion, Cabot and Kahl selected those cases which they consistently found produced good group discussions. In effect, the cases have been "pre-tested."

While the two volumes were designed as a unit, it is clearly evident that their use is not restricted in this respect. The cases can be employed with other concepts (or readings) than those presented in Volume I; and, of course, the concepts in Volume I can be used in any order or with a different set of cases. As the writers indicate, "the chapters and cases are so written that they are to a considerable extent complete in themselves, and such modifications can be readily made."

The number of colleges and universities in which integrated social science courses are being introduced into the undergraduate curriculum is on the increase. There is a desperate need for especially prepared written material for such a course for first-year college students. Both objectively reported case materials and succinct theoretical discussions for the undergraduate, lacking a background of specialized courses, are required. With respect to the former, Cabot and Kahl have made a definite contribution. Unfortunately, the same cannot be said with reference to the latter. While some portions of their conceptual discussions in Volume I probably can be used for integrated social science courses, on the whole much of it falls far short of the mark.

HAROLD PROSHANSKY
Brooklyn College

WHITING, JOHN W. M., & CHILD, IRVIN L. *Child training and personality: a cross-cultural study.* New Haven: Yale Univer. Press, 1953. Pp. vi+353. $5.00.

This is the report, within a specif-

ically defined theoretical framework, of a careful study in which 75 primitive societies are used as the individual variables. The sources of information about these societies are published anthropological reports, and are given in a classified list of references. The aim of the study is to test certain hypotheses about the effect of child-rearing practices on producing characteristic personality structures of the members of a culture. The authors have set up a series of hypotheses, and then tested them by applying to them available descriptions of the cultures concerned. These hypotheses are derived from a general behavior theory in which certain aspects of Freudian psychoanalytic theory are adapted to Hullian behavior theory.

Five systems of behavior are selected for study because they are considered basic to personality development, are subject to training and control in all cultures, and are treated in a wide variety of ways in the different cultures. These systems of behavior are: oral, anal, sexual, dependence, and aggression. For each culture the data about each of the systems were categorized according to (*a*) initial indulgence, (*b*) age of socialization, and (*c*) severity of socialization. Within these categories the classes of behaviors were defined in such a way as to maximize the objectivity of classifications, and to facilitate making useful distinctions. Three judges, working independently, rated and ranked the data.

The Freudian concept of fixation is considered basic, but it has been modified by division into positive and negative fixation. Evidence for both kinds of fixation is studied in each of the five systems. Positive fixation is expected to result from satisfactions experienced when drives in any behavior system are indulged.

Negative fixation should occur in a behavior system when training interferes with natural behavior tendencies. Several other concepts, such as "habit potential," "custom potential," and "custom complex," were defined as important in dealing with the cultural variables.

The personality characteristics studied are based on customs for which there was generally good information and for which, according to the authors' theory, there was expectation of a relation to the child-training practices studied. These characteristics are the attitudes in a society toward illness, its causes and cures, and toward death. Customs relating to illness were divided into: explanations of illness, performance therapies, avoidance therapies, and fear of others. Each of these was in turn subdivided according to the five basic systems of behavior. The data on illness from each culture were then classified by two judges who were kept in ignorance about the child-training classifications.

Interrelationships were studied among the various aspects of child-rearing practices, and between these and the personality characteristics of the adult members of the cultures, as evidenced in their customs in relation to illness. In many cases the theoretical hypotheses were confirmed or supported; in others the expected relationships were not found. The failure of verification of some of the hypotheses could be due to inadequacy of the data, which are often lacking in information pertinent to this inquiry. The findings are interesting and present a very challenging array of problems for future testing, by additional cultural studies and by experiment.

It is difficult to summarize the findings from the study, as they are so bound up with the context of in-terrelating variables and theories. But some examples may be given. The authors feel justified in their separation of positive and negative fixation. Negative fixation was most clearly demonstrated: it was shown in "the effects of anxiety developed during the period of socialization for each system of behavior." Positive fixation seemed to be related to a later period of childhood and to anal and sexual satisfactions. "Viewed as a whole, the evidence for a lasting effect of child training practices on personality is more striking and consistent for the three . . . systems— oral, dependence, and aggression." There was some evidence that guilt arises in the process of socialization as a consequence of identification with the punitive parent. In a very ingenious treatment of the data relating to the origins of unrealistic fear of other persons (paranoia) the authors "found that the fear of others is primarily associated with anxiety about aggression." Although the correlational method does not permit determination of the direction of the cause-effect relation, the authors feel they are justified in concluding that child-training practices do at least in part influence adult personality.

An incidental finding of interest is a comparison of child-training practices in midwestern urban middle-class families and the primitive cultures. By comparison, the Americans were exceptionally early and severe in toilet training (anal), and quite severe in weaning (oral) and in general socialization (average of five systems). Only in the socialization of aggression are they close to the norm for primitive cultures. However, the evaluation of these tendencies should be made in relation to their place in the total cultural pattern, and the authors explicitly make no recommendations regarding them.

This book has much to contribute, methodologically and theoretically, to the systematic treatment of the development of personality. But it is very difficult to read. The findings are buried and easily lost in the discussion and the tedious (though often important) qualifications of theoretical presentations. Perhaps this difficulty is unavoidable in working out this kind of complex general behavior theory. But it seems to this reviewer that a simpler style could make the theory a little more readily available and generally understood.

NANCY BAYLEY
University of California, Berkeley

ROSEN, JOHN N. *Direct analysis: selected papers.* New York: Grune & Stratton, 1953. Pp. vii + 184. $3.75.

The regard, respect, and human warmth for the mentally sick individual that John Rosen conveys in his public appearances are well reflected in the nine papers which have been collated into this slim volume. Eight of these papers have seen prior publication or presentation before psychiatric audiences, but the first paper, "Direct Analysis: General Principles," was specifically prepared for this book.

There are two basic premises from which Rosen proceeds: first, that unless physical pathology can be clearly demonstrated, psychoses are psychogenic in origin and are treatable by psychologic means; and second, that the nuclear core of the psychotic response is formed in the oral stage of infant dependence upon the mother. The effectiveness of direct analysis lies in the capacity of the therapist to enter without fear directly into the patient's psychosis and to clarify for him the dynamics of this oral deprivation by direct psychoanalytic interpretations as well as

by providing some of the absent love. Once reality contact has been firmly established, the therapy proceeds along more traditional psychoanalytic lines. On the basis of his experience, the author categorically asserts that he knows of no psychogenic psychosis that cannot be significantly improved by psychological therapy (i.e., psychoanalysis for Rosen). The author seems to have need for a rather uncritical, uncluttered, orthodox Freudian psychoanalytic theorizing to lend "scientific respectability" to his superb skill in interpersonal relationship therapy and to enable him to communicate to others what he believes to occur in the therapy situation. As an experimental method for studying the psychotic process, Rosen's techniques may prove extremely valuable. The limitations of application of this procedure to the massive problem of the psychotically ill are self-evident.

The interview and case material provide a fair amount of source material on the practice of direct analysis. The book is well worth reading not only for its contributions to the psychopathology of psychosis but also for its quality as a personal document.

M. ERIK WRIGHT
University of Kansas

VITELES, MORRIS S. *Motivation and morale in industry.* New York: Norton, 1953. Pp. xvi + 510. $9.50.

Viteles has undertaken, in this volume, to synthesize the data and reconcile differences from the ever increasing investigations of the "will-to-work" in industry. He has done an acceptable job in terms of the limitations of the studies available and the diverse orientation toward the problems of motivation in industry that the investigators on whose work he relies have. In doing this he

has produced a much needed review of the literature in this area for research workers and a handbook for personnel-minded members of management.

The book is directed primarily to members of management who are responsible for the development and administration of personnel policies. The emphasis throughout is on applications of research findings to problems of motivating employees. Viteles has developed a reasonably consistent set of principles for this purpose and has presented specific suggestions for their application. He also has pointed out inadequacies in available data and emphasized the need for further research.

In addition to affirming the importance of security, participation, status, and role in employee motivation, the author has re-examined the problem of economic motivation and has rescued it from the limbo to which it has been consigned by the Hawthorne investigation and certain attitude surveys. He has phrased properly the problem of financial incentives versus human relations in the motivation of workers: employees want both, and the problem confronting management is an integration of the two.

Viteles relies for his data on investigations classified roughly as "experimental" studies and attitude surveys. Typical of the former is the Hawthorne investigation. Data from attitude surveys represent a wide range of approaches, from those of Houser in the late twenties to the General Motors MJ Contest and work of the Michigan Survey Research Center and the National Industrial Conference Board. Most of the studies have appeared within the last five years. The investigations, for the most part, were conducted in industrial situations.

In spite of the excellent synthesis of these investigations, this reviewer is concerned with Viteles' failure to evaluate critically certain investigations on which he relies for his conclusions. It is true that he has an excellent chapter on the weakness of attitude survey methodology. Yet results of studies by the Opinion Research Corporation, Rose's study of workers' attitudes toward labor unions, and others are presented without making specific their limitations. Viteles cannot assume that the audience for which this volume is primarily intended, managerial personnel, will be well enough acquainted with these investigations or sufficiently sophisticated to evaluate their results without a specific statement from the author as to their limitations. In contrast, Viteles has been quite explicit as to the research weakness of the investigations dealing with group decision making in motivating employee behavior.

As indicated earlier, this volume is intended primarily for administrators concerned with personnel policy. This reviewer is somewhat apprehensive that the administrators who read this book will receive the idea that developing the will-to-work is primarily a matter of techniques rather than dependent on a fundamental philosophy of the dynamic relationship between a man and his job. There is a wide discrepancy between regarding "security" as a means of increasing worker productivity and regarding "security" as a way of living essential to the well-being of both the individual worker and his employer. The latter concept may be implicit in Viteles' treatment of incentives but it needs to be made explicit.

The author usually keeps in view the audience for which he is writing. The style is reasonably straightfor-

ward although the wealth of material cited at times results in the impression that the volume is an annotated bibliography. Statistical terminology at times will confuse the lay reader especially since no explanations of the concepts are offered.

Motivation and Morale in Industry has real significance for industrial psychology. It is the only comprehensive and compact source of bibliographical data in this area available to psychologists working on industrial problems. Even more important, it emphasizes the centrality of the problem of motivation in industrial psychology. It is significant that Viteles has chosen to write on this subject rather than to revise his *Industrial Psychology* published in 1932. It is an indication that psychologists concerned with industry are now moving away from a primary concern with the peripheral problems of employee selection, training, vision, and job analysis to the central problem of the work situation—the dynamics of behavior within the industrial organization.

WILLIAM McGEHEE
Fieldcrest Mills, Inc.

WALTER, W. GREY. *The living brain.* New York: Norton, 1953. Pp. 311. $3.95.

Starting with a delightfully written, simple story of the evolution of the nervous system and the emergence of the brain of man, Walter ends with some thoughts on the brain of tomorrow. He pays high tribute to Pavlov and Berger, whose contributions to conditioned reflexology and electroencephalography, respectively, are landmarks in our understanding of the brain. He attempts a convergence of the notions resident in these seemingly diverse areas. A brilliant thinker and writer, Walter does not hesitate to highlight his own role in the tremendous progress in electrophysiology in the last decade or so. For one looking on *de novo*, the impression might be gained that electroencephalography has been brought to its present state of development by the author singlehandedly.

The EEG and its elaborations by automatic analyzer and toposcope are looked upon as a suitable mirror for the brain's physiological activities, especially for detecting the significance of patterns which may correlate with the psychological phenomena of perception, learning, and personality. Certainly some interesting ideas are presented. The reviewer shares with Grey Walter the hopefulness of *rapprochement* in the realms of the physiological and psychological, but at best it will be a slow and laborious process. This is a book of ideas and hunches, flavored perhaps more with wish fulfillment than with empirical fact, but nevertheless an exciting account which psychologists might well read with profit.

Walter is greatly impressed with the subjective effects of flicker stimulation, and like others sees in photic driving a hopeful device for manipulating brain rhythms in frequency, phase, and space distribution with the goal of testing the hypothesis that brain rhythms serve a scanning function similar to that in television. Two chapters, six and seven, entitled "Learning about Learning" and "Steps from Chance to Meaning," ought to prove of interest and value to the psychologist, for here elements of neurophysiology, communication theory, feedback, and scanning mechanisms are brought to bear on the perceptive-learning-memory process. Much of Walter's thinking on such matters will seem somewhat more cavalier than comprehensive, and stems from analogical conceptions developed through

the construction of ingenious electro-mechanical models. The most elaborate of these is Machina Speculatrix, a bug-like toy, which has built into it potentialities for responsive behavior described as exploratory curiosity, sensory discrimination, negative and positive tropism, and other properties called discernment and self-recognition. With such a device he has identified what he believes are seven steps from chance to meaning. The circuit diagrams for a nerve model, M. speculatrix, and a conditioned reflex analogue are presented in appendices. All in all it is an interesting and provocative book.

DONALD B. LINDSLEY
University of California at Los Angeles

FESTINGER, LEON, & KATZ, DANIEL. (Eds.) *Research methods in the behavioral sciences.* New York: Dryden Press, 1953. Pp. xi+660. $5.90.

Because of the large number and variety of variables that may affect the phenomena they study, psychologists have been forced to pay particular attention to problems of research design and techniques of observation and measurement. *Research Methods in the Behavioral Sciences* is devoted to the consideration of these problems as they are faced by the investigator who is concerned with human behavior in its social setting.

The general attitude of the authors influences the organization and content of the text in at least two different ways. In the first place the methods discussed are conceived to apply widely to studies of social behavior regardless of the discipline—anthropology, psychology, sociology—to which an investigator formally claims allegiance. The emphasis is upon the type of question asked rather than upon the arbitrary division of the social sciences. It is for this reason

that the text has in its title the term "behavioral sciences." In the second place, the authors feel strongly that progress in the investigation of man's social behavior is most likely to be achieved by use of the basic logic of scientific methodology. This is the logic upon which research design in all sciences depends, hence, ". . . there are probably no social-psychological methods as such."

I would give an incorrect impression to imply that this text is primarily a discourse on the logic of scientific method illustrated with examples from research in the behavioral sciences. A distinction made by the authors between methods of research design and techniques by which a design may be put into effect is analogous to the military distinction between "strategy" and "tactics." Accepting that scientific logic is common to all sciences, the editors emphasize that " . . . specific techniques and approaches will vary, depending upon the subject matter . . ." under investigation. It is in their tactics that the behavioral sciences differ from other sciences.

The most important contribution of this text lies in the systematic manner in which the techniques available to the behavioral sciences are presented. The text follows the sequence of operations confronting the research worker from the time the question he is asking has been clearly stated to the time when he is ready to present a report on his efforts to answer it. A major portion of the book is devoted to procedures for sampling a population to be studied, and to techniques of collecting and analyzing data. In many cases techniques are described in the detail which is usually associated with a handbook. In this respect, the text is most useful to those who wish to check their own procedures as well as

those who are in the process of developing new research skills. The latter should understand that just because the text is so systematic in its organization does not mean that all methodological problems in social psychology have been solved.

It is understandable that, in attempting to be systematic, the authors have not always succeeded in introducing originality into their subject matter. However, because they have attempted to be systematic, they have included discussions of matters of importance to scientific method which have received little attention in similar texts in the past. With a bit more daring—or, perhaps, more space—they might have ventured to present more fully the logical bases of different patterns of research design. They perform a valuable service in discussing the basic assumptions and logical requirements that underlie the research techniques they are discussing. Problems of objective observation are considered in terms of the processes of hypothesis construction and of research design which precede them in the sequence of steps through which the investigator must pass. Systematic observation involves some form of measurement no matter how relatively crude it may be, and it is encouraging to find that a discussion of the theory of social measurement leads to an emphasis on the valuable roles which nonparametric methods of analysis may play in analyzing data of the type which most measuring instruments in the social sciences now provide.

The systematic coverage of the field which characterizes this text persists through the final chapter, where it is pointed out that the role of the social scientist frequently does not end with the completion of a research project. Questions of applying

research findings arise, bringing with them problems of a technological nature which are themselves in need of research. Since little knowledge regarding such problems is now available, the best that can be done is to discuss experiences in this branch of psychotechnology and to make a case for including such considerations in future research and training programs.

Research Methods in the Behavioral Sciences stands up well under the criticisms that are frequently leveled against texts to which a number of authors have contributed. The excellent integration of its thirteen chapters suggests that very careful and detailed planning must have preceded the actual writing. There is much less duplication of material than is usually found in such collaborative endeavors. Although the results of a Flesch count—old style or new—are not available, the general level of readability appears to be good and within the grasp of advanced undergraduate or graduate students.

For their present purposes the authors have drawn upon "... a limited pool of methodological research findings, to which they have themselves contributed." There is never a doubt that the authors have other than the same general orientation toward the field they are discussing. With few exceptions they are affiliated at present, or have been in the past, with the development of social psychology at the University of Michigan. This has advantages as far as the presentation of a coherent, well-integrated point of view is concerned, but in all probability it is a characteristic of the text which will bring comments from readers who would like to have seen material included that does not appear.

A lesson for some to learn from a

thoughtful appraisal of this text is the unsatisfactory nature of attempts to define a science in terms of the subject matter it investigates. Psychology can claim to be a science in so far as psychologists employ scientific methodology, based, as it is, upon certain well-established principles of reasoning conforming to certain strict requirements and accepting certain articles of faith. The authors have shown how the experimental and nonexperimental methods of science can be called to the aid of those who are interested in man's social behavior. This achievement will not come as a surprise to those who have followed the development of social psychology over the past number of years and know how forcefully and consistently Dr. Katz has pressed for careful attention to methodological problems. The present text illustrates that he and his associates have gone a long way in this direction, yet recognize that there is still a good distance to cover.

ROGER W. RUSSELL

University College, London

SARASON, SEYMOUR B. *Psychological problems in mental deficiency.* (2nd Ed.) New York: Harper, 1953. Pp. x+402. $5.00.

Dr. Sarason has written a very good, scholarly book. Professional students in the social and biological sciences and relatively well-informed laymen should find it unusually readable. Some readers may react against arguments by assertion and the sometimes polemic and hypercritical style, but these characteristics do not detract from the solid merits of the book. One gets the impression that it was written by an informed, critical psychologist who was stimulated by his experience to prod us out of the lethargy and pessimism which

are so frequently found among professional workers in the field of mental deficiency.

The book does not attempt a thorough coverage of the etiology, characteristics, and treatment of the great variety of pathologies that are thrown together in this so-called field. Rather, it is most valuable for its points of view, the discussion of research, and the questions raised. I see Sarason's book as one of the opening guns in a revolution which may overthrow or, hopefully, dismember this field of mental deficiency. For, if I read Sarason right or without too much projection, he is telling us that the mental defective is not what he seems or what the IQ says he is, that not all mentally defective children are *mentally* defective, and that many things can and should be done about their problems. In other words, the measured intellectual level is *not* the problem but is only a symptom. Sarason's emphases upon cultural problems and the effects of environmental stimulation and psychotherapy are important and sound.

The net result of Sarason's arguments should be the recognition that the field consists of many special pathologies, organic, genetic, and psychogenic, which should be diagnosed in terms of etiology and prognosis under appropriate treatment rather than in terms of intelligence test scores. He fails to make a complete break with the past, however, and the reader may get the impression that the term mental deficiency still somehow describes a meaningful category or homogeneous group.

The frequent use and implications of the term "garden variety defective" bother me. The term is used to refer both to cases of inherited deficiency and to psychogenic or cultur-

... determined cases. The term is ... chosen as it is not at all con... ... with Sarason's argument for ... ... individualized understanding and ... treatment, and it is rather meaning... ... less and offensive. Perhaps it is a ... ... to question such statements about projective techniques as "in recent years the psychological pen... dulum seems to have swung from the Binet to the Rorschach as the basis for a diagnosis of mental deficiency" (p 242). No evidence is given and Sarason's discussion would seem to contradict it.

The final chapters are concerned with the important problems of dealing with parents, the question of institutionalization, and the inadequacies of professional training. One can only applaud Sarason's position and treatment of these problems. On the whole, the book will undoubtedly improve the stature of professional work in this broad but neglected field.

KARL F. HEISER
The Vineland Training School

COMMITTEE ON COLORIMETRY OF THE OPTICAL SOCIETY OF AMERICA. *The science of color*. New York: Thomas Y. Crowell, 1953. Pp. xiii + 385. $7.00.

The principal contribution of this book may lie not so much in the factual data presented as in the exposition of "color" as a psychophysical concept. This concept is of such importance and broad application that it will undoubtedly have considerable influence on scientific procedures in other research fields of the special senses. Even if it took ten years to hammer out this basic idea, the results may compensate for an otherwise disappointing publication.

America's foremost specialists in color collaborated in producing this most authoritative book on color ... authoritative but inhomogeneous. Perhaps the work should have been presented as "The Present Status of Color as Viewed by Twenty-three Specialists" rather than as a definitive work. It is still an interim report, a combination of textbook, handbook, essay, and committee paper. It covers the principal phases of color: physical, psychophysical, and psychological; it gives a miniature history of optics, a one-reel travelogue through ancient civilizations, and the usual anatomical digest. The rapid run-through of instruments, methods, and standards is given with an economy suggesting that the bell is about to ring and the students are rustling their notebooks. As a book, it is marked by the lack of an over-all plan, unevenness, discontinuity, minor errors, and many irritating kinks in arrangement.

Nevertheless, it will stand on the right-hand bookshelf for many years because it has features which will not soon be supplanted. The references are comprehensive and well selected, especially in the psychological field. Professional workers will appreciate the most complete assortment of colorimetric tables that has yet appeared in one volume. Its major offering, "The Concept of Color," is thoroughly expounded for the reader who is willing to make the effort, though the material is scattered in the Introduction and parts of Chapters 2 and 7. The diagram in Figure 5 is invaluable to this explanation and might well have been used as the frontispiece of the book. The glossary-index will be most useful to visitors from other scientific fields. Excellent color plates help to clarify the psychological organization of color sensations into three dimen-

sions. The large, diffuse field of color phenomena has been painstakingly screened and organized in the two chapters on "Psychological Concepts," though it is very difficult for the reader to find his way through such diverse material without a special table of contents.

If the book has failed to reach its goal, we know it is not for lack of prodigious effort on the part of a number of the editors, for lack of erudition, or for lack of physical resources. Perhaps the difficulty lies in the times. Our best scientists are pushed at top speed on their jobs, and there is less time than formerly for the repose in which to reflect on the broader aspects of a part-time, unpaid assignment which must be taken out of family and social life. Also, all scientific fields have expanded so rapidly that it has become increasingly difficult for specialists to encompass them editorially. It is interesting to consider what form the book might have if the editor of the *Scientific Monthly*, for example, were asked to plan and supervise the next edition.

It is curious that a group of scientists with international scope should write with such an insular viewpoint. This is strictly "the American science of color." There is no suggestion here of the enlightening theoretical approach to color space developed by Schrodinger and Bouma, or of current developments on the Continent. Nor would the reader suspect the major contributions which have been made in psychophysical research methods by the British in their several institutions.

The neglect of the British field of research emphasizes an omission of fundamental importance in this as well as in most other recent books on color. Much space is devoted to methods, formulae, tables, and statistics for colorimetric work; but scarcely a paragraph can be found on the corresponding methodology and tools of experimental psychology. The superb instrumentation which has been developed by the physical and electronic technician has made it possible for anyone with a few thousand dollars to lay claim to be a "colorist." This can easily lead to the dangerously naive viewpoint that psychological methodology is unnecessary or perhaps may be expected to "just come naturally."

A book which is designed to fill such an important need deserves a future edition in which the psychophysical and psychological sections are strengthened to balance the physical, in which the pertinency of other sections is reviewed, the inclusion of now-omitted sciences proposed—and with the whole plan supervised by a professional editor-in-charge.

Dean Farnsworth
Visual Engineering Section,
U. S. Naval Submarine Base,
New London

Moreno, J. L. *Who shall survive?* New York: Beacon House, 1953. Pp. cxiv+763. $10.00.

This is a very large book. It has over 100 pages of preface before it gets to page 1. Afterwards it still runs to over 700 pages. In the long preface the author tells us that:

1. Sociometry is a religion as well as a science (e.g., page xv): "I tried to do through sociometry what religion without science has failed to accomplish in the past. . . . "

2. The author is a genius (e.g., page xxxvii): "All my books (nine) published between 1919 and 1925, were anonymous. The natural state of genius is anonymity . . . *The*

Words of the Father I wrote with red ink on the walls of an Austrian castle."

3. Sociometry is the final development (e.g., page lxvii): "Actually, I have written two bibles, an old testament and a new testament, *The Words of the Father* and *Who Shall Survive?*"

4. Psychoanalysis is finished (e.g., page liv): " . . . the psychoanalytic system was stillborn to start with. . . . The psychoanalytic couch has become a piece of furniture in the sociodynamic field of the psychodramatic stage."

5. Group therapy, group dynamics, and, it seems, all work on small groups and large groups are developments of, or secessions from, the sociometric movement (pages lxix, lxx, and xcvii to cviii).

The body of the text is devoted to an exposition of the theory, methods, and accomplishments of sociometry. While it is difficult to give a brief review of over 700 pages, I will try to list the major points which are made.

1. The most important concepts of the sociometric system are spontaneity, creativity, Tele, the social atom, and underlying social networks. There is considerable discussion which conveys to the reader a vague feeling concerning the meaning of these terms. There are no very specific or rigorous definitions.

2. The basic techniques of sociometry are psychodrama, sociodrama, role-playing, group psychotherapy, and the sociometric test. Only the discussion of the last of these is at all elaborated in detail. There is little more than brief mention of the others. The only empirical data presented concern the sociometric test, the essence of which is asking people with whom they want to associate.

3. A major portion is devoted to detailed description of two studies using this sociometric test. Data are presented to show that the distribution of choices is different from chance, the pattern of choices changes with age, attractions among persons are not always in accordance with formal or existing associations, and that if groups are made up in accordance with the desires of the members, these groups hold the members' interest better.

To this reviewer it seems that there are many pages devoted to stating how important sociometry is. There is, however, little space devoted to any convincing empirical demonstration that these claims are valid.

LEON FESTINGER.
University of Minnesota.

MORSE, NANCY C. *Satisfaction in the white-collar job.* Ann Arbor: Univer. of Michigan, 1953. Pp. viii + 235. $3.50.

It has long been part of the folklore of industry that a satisfied employee is a productive employee. Recent investigations of the Michigan Survey Research Center have questioned this assumption. The indication that there is no one-to-one relationship between satisfaction and productivity has been disturbing both theoretically and practically. Morse has attempted in this volume to reduce theoretical tension by theory construction.

Satisfaction, according to Morse, is "the individual's perception of the tension-level within him" (p. 32). This tension-level in turn is a function of what an individual wants and what he receives (environmental return). Morse further considers the amount of satisfaction derived from a given activity as a separate variable from motivation, i.e., "the willing-

ness to expend energy in a given activity" (p. 11).

Productivity is, however, a much more complex phenomenon to predict than satisfaction, as it is dependent on more factors. Morse postulates at least seven. The only factor which productivity has in common with satisfaction is the strength of needs. Accordingly, individuals will experience high satisfaction and be high in productivity if they have strong needs which are "predictably" related to productivity and where environmental return from the job is high. If environmental return is low, productivity may be high but the amount of satisfaction will be low.

Morse emphasizes throughout the volume that the hypotheses concerning the nature of satisfaction and of productivity and their relationship are tentative and will require further investigation to verify, modify, or refute. As an indication of their possible validity she offers evidence from the analysis of interview data with 742 clerical employees and 73 first- and second-line supervisors in a large corporation. She points out the limitations, and possibly atypical nature, of this population.

Her evidence, as far as it goes, supports her hypotheses concerning both satisfaction and productivity.. The limitations of her data, in addition to population limitations, seem to be that the investigation was not originally planned to test hypotheses concerning these two variables. For example, her indices of job satisfaction are limited to only three measures. It seems that data collected for some other purpose were analyzed subsequently to throw light on the hypotheses presented.

She offers no evidence for the statement that the amount of satisfaction is a separate variable from motivation. In fact, her definition of motivation as well as of morale differs from the ones usually found in texts on the subject. An author has the privilege of defining the terms. But unless this serves some useful purpose, the use of familiar terms in a different context tends to obscure rather than clarify meanings.

Her evidence suffers from the weaknesses inherent in the interview method of securing data on problems of motivation, morale, and satisfaction. This is acknowledged tacitly in the discussion of the differences found between employees who receive general supervision as contrasted with closely supervised employees. The former were less enthusiastic about working for the company. This could have arisen, according to Morse, because "they maybe are in a situation which demands less strict conformity, they are more free to express reservations and criticisms" (p. 142). It seems that for a critical test of her hypotheses, interview data will have to be supplemented with data collected by other means.

Morse has, however, taken time out from collecting data to develop theory. It is the opinion of this reviewer that too little theory development has characterized the investigations psychologists have made in industry. The result has been a collection of unrelated and frequently contradictory data. A real need for advances in industrial psychology is good theory making. Morse has taken a step in the right direction.

WILLIAM McGEHEE.
Fieldcrest Mills, Inc.

BACH, GEORGE R. *Intensive group psychotherapy.* New York: Ronald, 1954. Pp. xi + 446. $6.00

BLACKHURST, J. HERBERT. *Body-mind and creativity.* New York: Philosophical Library, 1954. Pp. 186. $3.00.

BURROW, TRIGANT. *Science and man's behavior; the contribution of phylobiology.* (William E. Galt, Ed.) New York: Philosophical Library, 1953. Pp. xii + 564. $6.00.

CHRISTIE, RICHARD, & JAHODA, MARIE. (Eds.) *Studies in the scope and method of the authoritarian personality.* Glencoe: The Free Press, 1954. Pp. 279. $4.50.

ELLIS, ALBERT. *The American sexual tragedy.* New York: Twayne, 1954. Pp. 288. $4.50.

ELLIS, ALBERT. (Ed.) *Sex life of the American woman and the Kinsey report.* New York: Greenberg, 1954. Pp. 214. $2.75.

EYSENCK, H. J. *Uses and abuses of psychology.* Baltimore: Penguin, 1953. Pp. 318. $.65.

FRENCH, THOMAS M. *The integration of behavior.* Vol. II. *The integrative process in dreams.* Chicago: Univer. of Chicago Press, 1954. Pp. xi + 367. $6.50.

GANZ, MADELAINE. *The psychology of Alfred Adler and the development of the child.* (Trans. by Philip Mairet.) New York: Humanities Press, 1953. (First published in Geneva, 1935.) Pp. xxiii + 203. $4.50.

GRAY, J. STANLEY. (Ed.) *Psychology applied to human affairs.* (2nd Ed.) New York: McGraw-Hill, 1954. Pp. vii + 581. $6.00.

HAMMOND, KENNETH R., & ALLEN,

JEREMIAH M. *Writing clinical reports.* New York: Prentice-Hall, 1953. Pp. xii + 235. $5.35.

HARDING, D. W. *Social psychology and individual values.* London: Hutchinson, 1953. Pp. vii + 184. $1.80 text ed., $2.40 trade ed.

INGHAM, HARRINGTON V., & LOVE, LEONORE R. *The process of psychotherapy.* New York: McGraw-Hill, 1954. Pp. ix + 270. $5.00.

KLOPFER, BRUNO, AINSWORTH, MARY D., KLOPFER, WALTER G., & HOLT, ROBERT R. *Developments in the Rorschach technique.* Yonkers, N. Y.: World Book Co., 1954. Pp. x + 726.

LANDIS, CARNEY. *An annotated bibliography of flicker fusion phenomena* (covering the period 1740–1952). Ann Arbor: Armed Forces National Research Council, 1953. Pp. vi + 130.

LIBO, LESTER M. *Measuring group cohesiveness.* Ann Arbor: Institute for Social Research, Univer. of Michigan, 1953. Pp. ix + 111. $2.00.

McCURDY, HAROLD GRIER. *The personality of Shakespeare; a venture in psychological method.* New Haven: Yale Univer. Press, 1953. Pp. xi + 243. $5.00.

MARCUSE, F. L. (Ed.) *Areas of psychology.* New York: Harper, 1954. Pp. viii + 532. $5.00.

MOLONEY, JAMES CLARK. *Understanding the Japanese mind.* New York: Philosophical Library, 1954. Pp. xviii + 252. $3.50.

PERRY, RALPH BARTON. *Realms of value; a critique of human civilization.* Cambridge: Harvard Univer. Press, 1954. Pp. xii + 497. $7.50.

PODOLSKY, EDWARD. (Ed.) *Music*

therapy. New York: Philosophical Library, 1954. Pp. xii+335. $6.00.

REMMERS, H. H. *Introduction to* ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ New York: Harper, 1954. Pp. viii+437. $5.00.

SAPPENFIELD, BERT R. *Personality dynamics; an integrative psychology of adjustment.* New York: Knopf, 1954. Pp. xiv+428. $5.50.

STOLZ, LOIS MEEK, and collaborators. *Father relations of war-born children.* Stanford: Stanford Univer. Press, 1954. Pp. viii+365. $4.00.

STONE, CALVIN P. (Ed.) *Annual review of psychology.* Stanford: Annual Reviews, 1954. Pp. ix+448. $7.00.

STOTT, D. H. *Saving children from delinquency.* New York: Philosophical Library, 1953. Pp. x+266. $4.75.

SWARTZ, HARRY. *The allergic child.* New York: Coward-McCann, 1954. Pp. xvii+297. $3.95.

TAYLOR, W. S. *Dynamic and abnormal psychology.* New York: American Book Co., 1954. Pp. xx+658. $5.50.

VEDDER, CLYDE B. (Ed.) *The juvenile offender; perspective and readings.* Garden City: Doubleday, 1954. Pp. xii+510. $6.00.

VICTOROFF, DAVID. *G. H. Mead sociologue et philosophe.* Paris: Presses Universitaires de France, 1953. Pp. 150. 600 fr.

WEVER, ERNEST GLEN & LAWRENCE MERLE. *Physiological acoustics.* Princeton: Princeton Univer. Press, 1954. Pp. xii+454. $10.00.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954. Pp. xxvi+571. $7.50.

WOLFFHEIM, NELLY. *Psychology in the nursery school.* New York: Philosophical Library, 1953. Pp. 144. $3.75.

Psychological Bulletin

THE CRITICAL INCIDENT TECHNIQUE

JOHN C. FLANAGAN
American Institute for Research and University of Pittsburgh

During the past ten years the writer and various collaborators have been engaged in developing and utilizing a method that has been named the "critical incident technique." It is the purpose of this article to describe the development of this methodology, its fundamental principles, and its present status. In addition, the findings of a considerable number of studies making use of the critical incident technique will be briefly reviewed and certain possible further uses of the technique will be indicated.

The critical incident technique consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles. The critical incident technique outlines procedures for collecting observed incidents having special significance and meeting systematically defined criteria.

By an incident is meant any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act. To be critical, an incident must occur in a situation where the purpose or intent of the act seems fairly clear to the observer and where its consequences are sufficiently definite to leave little doubt concerning its effects.

Certainly in its broad outlines and basic approach the critical incident technique has very little which is new about it. People have been making observations on other people for centuries. The work of many of the great writers of the past indicates that they were keen observers of their fellow men. Some of these writers must have relied on detailed notes made from their observations. Others may have had unusual abilities to reconstruct memory images in vivid detail. Some may have even made a series of relatively systematic observations on many instances of a particular type of behavior. Perhaps what is most conspicuously needed to supplement these activities is a set of procedures for analyzing and synthesizing such observations into a number of relationships that can be tested by making additional observations under more carefully controlled conditions.

BACKGROUND AND EARLY DEVELOPMENTS

The roots of the present procedures can be traced back directly to the studies of Sir Francis Galton nearly 70 years ago, and to later developments such as time sampling studies of recreational activities, controlled observation tests, and anecdotal rec-

ords. The critical incident technique as such, however, can best be regarded as an outgrowth of studies in the Aviation Psychology Program of the United States Army Air Forces in World War II. The Aviation Psychology Program was established in the summer of 1941 to develop procedures for the selection and classification of aircrews.

One of the first studies (40) carried out in this program was the analysis of the specific reasons for failure in learning to fly that were reported for 1,000 pilot candidates eliminated from flight training schools in the summer and early fall of 1941. The basic source used in this analysis was the proceedings of the elimination boards. In these proceedings the pilot instructors and check pilots reported their reasons for eliminating the particular pilot. It was found that many of the reasons given were clichés and stereotypes such as "lack of inherent flying ability" and "inadequate sense of sustentation," or generalizations such as "unsuitable temperament," "poor judgment," or "insufficient progress." However, along with these a number of specific observations of particular behaviors were reported. This study provided the basis for the research program on selecting pilots. Although it was found very useful, it also indicated very clearly the need for better procedures for obtaining a representative sample of factual incidents regarding pilot performance.

A second study (13), which emphasized the importance of factual reports on performance made by competent observers, was carried out in the winter of 1943–1944 in the 8th, 9th, 12th, and 15th Air Forces. This study collected the reasons for the failures of bombing missions as reported in the Group Mission Reports.

Although in the preparation of these reports much greater emphasis was given to determining the precise facts in the case, it was apparent that in many instances the official reports did not provide a complete record of all the important events. Even with these limitations, the information given was found to be of considerable value, and the systematic tabulations that were prepared provided the basis for a series of recommendations that resulted in important changes in Air Force selection and training procedures.

In the summer of 1944 a series of studies (74) was planned on the problem of combat leadership in the United States Army Air Forces. These represent the first large-scale, systematic effort to gather specific incidents of effective or ineffective behavior with respect to a designated activity. The instructions asked the combat veterans to report incidents observed by them that involved behavior which was especially helpful or inadequate in accomplishing the assigned mission. The statement finished with the request, "Describe the officer's action. What did he do?" Several thousand incidents were collected in this way and analyzed to provide a relatively objective and factual definition of effective combat leadership. The resulting set of descriptive categories was called the "critical requirements" of combat leadership.

Another study (74) conducted in the Aviation Psychology Program involved a survey of disorientation while flying.[1] Disorientation in this

[1] This study was planned by Paul M. Fitts, Jr., who also contributed to the previously mentioned USAAF studies and planned and carried out the interview study with pilots described below on the design of instruments, controls, and arrangements.

study was defined to include any experience denoting uncertainty as to one's spatial position in relation to the vertical. In this study pilots returning from combat were asked "to think of some occasion during combat flying in which you personally experienced feelings of acute disorientation or strong vertigo." They were then asked to describe what they "saw, heard, or felt that brought on the experience." This study led to a number of recommendations regarding changes in cockpit and instrument panel design and in training in order to overcome and prevent vertigo while flying.

In a project carried out in the Aviation Psychology Program in 1946, Fitts and Jones (12) collected descriptions of specific experiences from pilots in taking off, flying on instruments, landing, using controls, and using instruments. These interviews with pilots were electrically recorded. They provided many factual incidents that were used as a basis for planning research on the design of instruments and controls and the arrangement of these within the cockpit.

In addition to the collection of specific incidents and the formulation of critical requirements, as outlined above, the summary volume (13) for the Aviation Psychology Program Research Reports contained a discussion of the theoretical basis of procedures for obtaining the critical requirements of a particular activity. Perhaps the best method of describing the status of these procedures at the close of the war is to quote from the discussion in this summary volume, which was written in the late spring of 1946. In the section on techniques for defining job requirements, the present author wrote as follows:

The principal objective of job analysis procedures should be the determination of critical requirements. These requirements include those which have been demonstrated to have made the difference between success and failure in carrying out an important part of the job assigned in a significant number of instances. Too often, statements regarding job requirements are merely lists of all the desirable traits of human beings. These are practically no help in selecting, classifying, or training individuals for specific jobs. To obtain valid information regarding the truly critical requirements for success in a specific assignment, procedures were developed in the Aviation Psychology Program for making systematic analyses of causes of good and poor performance.

Essentially, the procedure was to obtain first-hand reports, or reports from objective records, of satisfactory and unsatisfactory execution of the task assigned. The cooperating individual described a situation in which success or failure was determined by specific reported causes.

This procedure was found very effective in obtaining information from individuals concerning their own errors, from subordinates concerning errors of their superiors, from supervisors with respect to their subordinates, and also from participants with respect to co-participants (13, pp. 273–274).

DEVELOPMENTAL STUDIES AT THE AMERICAN INSTITUTE FOR RESEARCH

At the close of World War II some of the psychologists who had participated in the USAAF Aviation Psychology Program established the American Institute for Research, a nonprofit scientific and educational organization. The aim of this organization is the systematic study of human behavior through a coordinated program of scientific research that follows the same general principles developed in the Aviation Psychology Program. It was in connection with the first two studies undertaken by the Institute in the spring of 1947 that the critical incident technique was more formally developed and given its present name. These studies were natural exten-

sions of the previous research in the Aviation Psychology Program. The study reported by Preston (52) dealt with the determination of the critical requirements for the work of an officer in the United States Air Force. In this study, many of the procedural problems were first subjected to systematic tryout and evaluation. Six hundred and forty officers were interviewed, and a total of 3,029 critical incidents were obtained. This led to the development of a set of 58 critical requirements classified into six major areas. The second study, reported by Gordon (27, 28), was carried out to determine the critical requirements of a commercial airline pilot. In this study, several different sources were used to establish the critical requirements of the airline pilot. These included training records, flight check records including the specific comments of check pilots, critical pilot behaviors reported in accident records, and critical incidents reported anonymously in interviews by the pilots themselves. From this study, 733 critical pilot behaviors were classified into 24 critical requirements of the airline pilot's job. These were used to develop selection tests to measure the aptitudes and other personality characteristics found critical for success in the job. They also provided the basic data for the formulation of an objective flight check to determine the eligibility of applicants for the airline transport rating.

The third application of the critical incident technique by the staff of the American Institute for Research was in obtaining the critical requirements for research personnel on a project sponsored by the Psychological Sciences Division of the Office of Naval Research. In this study (20), about 500 scientists in 20 research laboratories were interviewed. These scientists reported more than 2,500 critical incidents. The critical behaviors were used to formulate inductively a set of 36 categories, which constitutes the critical requirements for the effective performance of the duties of research personnel in the physical sciences. This initial study provided the basis for the development of selection tests, proficiency measures, and procedures for evaluating both job performance and the research report.

Another project undertaken by the American Institute for Research in the spring of 1948 provided valuable experience with the critical incident technique. This study, reported by Nagay (48), was done for the Civil Aeronautics Administration under the sponsorship of the Committee on Aviation Psychology of the National Research Council. It was concerned with the air route traffic controller's job. One of the innovations in this study was the use of personnel of the Civil Aeronautics Administration who had no previous psychological training in collecting critical incidents by means of personal interviews. In previous studies all such interviewing had been conducted by psychologists with extensive training in such procedures. In this study, aeronautical specialists from each of the seven regions conducted the interviews in their regions after a brief training period. An interesting finding from this study was the clear reflection of seasonal variations in flying conditions in the types of incidents reported. The study also demonstrated the selective recall of dramatic or other special types of incidents. This bias was especially noticeable in the incidents reported several months after their occurrence. The incidents obtained in this

study were used to develop procedures for evaluating the proficiency of air route traffic controllers and also for developing a battery of selection tests for this type of personnel.

In the spring of 1949 the American Institute for Research undertook a study to determine the critical job requirements for the hourly wage employees in the Delco-Remy Division of the General Motors Corporation. This study, reported by Miller and Flanagan (46), was the first application of these techniques in an industrial situation. Foremen who were members of a committee appointed to develop employee evaluation procedures collected 2,500 critical incidents in interviews with the other foremen in the plants. On the basis of these data a form was prepared for collecting incidents on a day-to-day basis as a continuous record of job performance.

Using this form, the Performance Record for Hourly Wage Employees (21), three groups of foremen kept records on the performance of their employees for a two-week period. A group of 24 foremen recorded incidents daily; another group of 24 foremen reported incidents at the end of each week; and a third group containing the same number of foremen reported incidents only at the end of the two-week period. The three groups of foremen represented comparable conditions of work and supervision. The foremen reporting daily reported 315 critical incidents; the foremen reporting weekly, 155 incidents; and the foremen reporting only once at the end of two weeks reported 63 incidents. Thus, foremen who reported only at the end of the week had forgotten approximately one half of the incidents they would have reported under a daily reporting plan. The foremen who reported

only at the end of the two-week period appeared to have forgotten 80 per cent of the incidents observed. Although it is possible that the findings may be partially attributed to the fact that the foremen making daily records actually observed more critical incidents because of the daily reminder at the time of recording, it is clear that much better results can be expected when daily recording is used.

Another analysis based on data collected at the Delco-Remy Division compared the number of critical incidents of various types obtained from interviews with those recorded daily by the foremen on the performance record. Although there were some differences in the relative frequencies for specific categories, the general patterns appeared to be quite similar. These results suggest that critical incidents obtained from interviews can be relied on to provide a relatively accurate account of job performance if suitable precautions are taken to prevent systematic bias.

In addition to the development of the performance record described above, the critical incidents collected in this study were used as the basis for constructing selection tests covering both aptitude (18) and attitude (2) factors.

STUDIES CARRIED OUT AT THE UNIVERSITY OF PITTSBURGH

A substantial number of studies have been carried out in the department of psychology at the University of Pittsburgh by students working for advanced degrees under the author's direction. Most of these studies had as their objective the determination of the critical requirements for a specific occupational group or activity. Many of them also included contributions to technique. In 1949

Wagner (66) completed a dissertation on the critical requirements for dentists. In this study, critical incidents were obtained from three sources: patients, dentists, and dental school instructors. The incidents were classified into four main aspects of the dentist's job: (a) demonstrating technical proficiency; (b) handling patient relationships; (c) accepting professional responsibility; and (d) accepting personal responsibility. As might be expected, the patients did not report as large a proportion of incidents for demonstrating technical proficiency or accepting professional responsibility as did the other two groups, and the instructors reported only a relatively small proportion of their incidents in the area of handling patient relationships.

On the basis of the findings from this study, a battery of selection tests was developed for use by the University of Pittsburgh School of Dentistry. A number of proficiency tests for measuring ability with respect to certain of the critical requirements were also developed using these results as a basis.

Another dissertation completed in 1949 was Finkle's (11) study of the critical requirements of industrial foremen. This study was conducted in the East Pittsburgh plant of the Westinghouse Electric Corporation. Critical incidents were obtained from foremen, general foremen, and staff personnel. A number of points pertaining to technique were studied.

One finding was in reference to the effect on the types of incidents obtained of the degree of importance or exceptionalness set up as a criterion for reporting or ignoring incidents. The incidents obtained from the use of questions that asked for incidents only slightly removed from the norm were compared with incidents obtained from questions intended to elicit more definitely effective or ineffective behaviors. Some examples of these questions are:

1. Think of a time when a foreman has done something that you felt should be commended because it seemed to be in your opinion an example of good foremanship. (Effective—slight deviation from norm.)

2. Think of a time when a foreman did something that you thought was not up to par. (Ineffective—slight deviation from norm.)

3. Think of a time when a foreman has, in your opinion, shown definitely good foremanship—the type of action that points out the superior foreman. (Effective—substantial deviation from the norm.)

4. Think of a time when a foreman has, in your opinion, shown poor foremanship—the sort of action which if repeated would indicate that the man was not an effective foreman. (Ineffective—substantial deviation from norm.)

The frequencies of incidents obtained in each of the 40 categories into which the effective behaviors were classified were compared for the questions requesting slight and substantial deviations from the norm, and the significance of the differences was tested by means of the chi-square test. Two of the differences were significant at the 1 per cent level and one at the 5 per cent level. Comparisons of the frequencies in each of the 40 categories for ineffective incidents failed to reveal any chi squares significant at either the 5 per cent or the 1 per cent level.

The questions involving only a slight deviation from the norm resulted in more effective incidents concerned with gaining the respect and loyalty of the workers and also in more incidents that involved making, encouraging, and accepting suggestions. They produced significantly fewer incidents regarding fitting men to jobs. The small number of significant differences—only three in 80 comparisons—suggests that the

..... of incidents obtained are not very greatly changed by variations in wording of the questions comparable to those shown above. It seems likely that this is at least partially due to the fact that the persons interviewed report only incidents that represent a fairly substantial deviation from the norm regardless of the precise wording of the question asked.

Another comparison made in this study related to the influence of asking for an effective or an ineffective incident first. About 10 per cent more incidents were obtained from booklets requesting effective incidents first than from booklets requesting ineffective incidents first. This difference was sufficiently small so that it could reasonably be attributed to chance sampling fluctuations.

The incidents collected in this study were used, along with other data, in the preparation of a Performance Record for Foremen and Supervisors (23).

A study was conducted by Nevins (50) on the critical requirements of bookkeepers in sales companies. She collected incidents relating to applicants for bookkeeping positions as well as for employees working in this capacity.

For the collection of the information about the practicing bookkeepers, a modification in the critical incident technique was made. This was done because, in the bookkeeping profession, success and failure are usually defined in terms of persistent behavioral patterns. Occasional mistakes in adding and balancing accounts are expected, but repeated errors are considered serious. Instead of the single incident, therefore, many of the items included represented either a pattern of behaviors or a series of similar behaviors.

Weislogel (72) determined the critical requirements for life insurance agency heads. A principal feature of his study related to the comparison of two types of agency heads—managers and general agents. It was believed that the critical behaviors for one type of agency head might provide a different pattern than that obtained for the other. This hypothesis was not confirmed by the analysis of the obtained incidents. The patterns of critical requirements were found to be quite similar for the two types of administrators.

Smit (58) carried out a study to determine the critical requirements for instructors of general psychology courses. Perhaps the finding of most general importance in this study was the existence of substantial differences between the patterns of critical incidents reported by students and faculty. The faculty reported a significantly larger percentage of effective behaviors in the following areas: giving demonstrations or experiments, using discussion group techniques, encouraging and ascertaining students' ideas and opinions.

The students, on the other hand, contributed a larger percentage of behaviors in the following areas: reviewing examinations, distributing grades, and explaining grades; using lecture aids such as drawings, charts, movies, models, and apparatus; using project techniques; giving test questions on assigned material; helping students after class and during class recess; the manner of the instructor.

The faculty reported a larger percentage of ineffective behaviors concerning maintaining order. The ineffective behaviors that were reported in a larger percentage by students involved these areas: presenting requirements of the course, using effective methods of expression, dealing

with students' questions, pointing out fallacies, reviewing and summarizing basic facts and principles, using project techniques, using verbal diagnostic teaching techniques, achievement testing students on assigned material, objective type achievement testing, using humor.

This is a good illustration of the problem of the competence of various types of available observers to evaluate the contribution to the general aim of the activity of a specific action. Examination of the reports from students indicated a somewhat limited sphere of competence. Apparently one of the principal reasons for this was the lack of perspective on the part of the students and their inability to keep the general aim of the instructor clearly in mind because of its divergence from their own immediate aims. In many cases, this latter aim seemed to be directed toward achieving a satisfactory grade in the course.

Eilbert (7) developed a functional description of emotional immaturity. The contributors of critical incidents included psychiatrists, psychologists, psychiatric social workers, occupational therapists, nurses, and corpsmen from a military hospital, plus 13 psychologists in nonmilitary organizations. The subjects of the incidents were primarily patients under psychiatric care.

The contributors were given a form that oriented them to the concept "emotional immaturity" by suggesting that it was revealed generally by childlike modes of behavior. The questions used to elicit incidents were: Have you recently thought of someone as being emotionally immature (regardless of diagnosis)? What specifically happened that gave you this impression? What would have been a more mature reaction to the same situation?

Because of the indefinite nature of the concept, it was felt that a check should be made on the contributor's understanding of his task. Twenty of the participating persons were asked to summarize briefly their interpretation of what they had been asked to do. This appeared to be very useful in developing the phrasing of the questions so that they were uniformly interpreted by the various observers.

The author of the study classified all the immaturities on the basis of a classification system developed from preliminary categorizations prepared by six of the contributors. This classification was submitted to 14 psychiatrists for review. They were asked to indicate which of the categories they were willing to accept as a type of immaturity as the term had been defined in an official document. More than half the categories were accepted by at least 13 of the 14 judges, and none was rejected by more than 50 per cent of the judges. It was felt then that the system was acceptable.

This study illustrates the application of the critical incident technique to the study of personality. It is believed that this study provides an excellent example of the possibilities for developing more specific behavioral descriptions.

Folley (24) reported on the critical requirements of sales clerks in department stores. The behaviors were abstracted from narrative records of individual shopping incidents written by shoppers who were relatively inexperienced in evaluating sales personnel. For various reasons, including the competence of the observers, their training, and their limited point of view, the resulting description must be regarded as only partial.

In the past few years, many other individuals and groups have made use of the techniques described

above, or modifications of them, in a wide variety of studies. Some of these studies on which reports are being published will be reviewed briefly in the section on applications.

THE PROCEDURE IN ITS PRESENT FORM

From the foregoing discussion, it is clear that the critical incident technique is essentially a procedure for gathering certain important facts concerning behavior in defined situations. It should be emphasized that the critical incident technique does not consist of a single rigid set of rules governing such data collection. Rather it should be thought of as a flexible set of principles which must be modified and adapted to meet the specific situation at hand.

The essence of the technique is that only simple types of judgments are required of the observer, reports from only qualified observers are included, and all observations are evaluated by the observer in terms of an agreed upon statement of the purpose of the activity. Of course, simplicity of judgments is a relative matter. The extent to which a reported observation can be accepted as a fact depends primarily on the objectivity of this observation. By objectivity is meant the tendency for a number of independent observers to make the same report. Judgments that two things have the same effect or that one has more or less effect than the other with respect to some defined purpose or goal represent the simplest types of judgments that can be made. The accuracy and therefore the objectivity of the judgments depend on the precision with which the characteristic has been defined and the competence of the observer in interpreting this definition with relation to the incident observed. In this latter process, certain more difficult

types of judgments are required regarding the relevance of various conditions and actions on the observed success in attaining the defined purpose for this activity.

It is believed that a fair degree of success has been achieved in developing procedures that will be of assistance in gathering facts in a rather objective fashion with only a minimum of inferences and interpretations of a more subjective nature. With respect to two other steps that are essential if these incidents are to be of value a comparable degree of objectivity has not yet been obtained. In both instances, the subjective factors seem clearly due to current deficiencies in psychological knowledge.

The first of these two other steps consists of the classification of the critical incidents. In the absence of an adequate theory of human behavior, this step is usually an inductive one and is relatively subjective. Once a classification system has been developed for any given type of critical incidents, a fairly satisfactory degree of objectivity can be achieved in placing the incidents in the defined categories.

The second step refers to inferences regarding practical procedures for improving performance based on the observed incidents. Again, in our present stage of psychological knowledge, we are rarely able to deduce or predict with a high degree of confidence the effects of specific selection, training, or operating procedures on future behaviors of the type observed. The incidents must be studied in the light of relevant established principles of human behavior and of the known facts regarding background factors and conditions operating in the specific situation. From this total picture hypotheses are formulated. In only a few types of activities are there

both sufficient established principles and sufficient information regarding the effective factors in the situation to provide a high degree of confidence in the resulting hypotheses regarding specific procedures for improving the effectiveness of the results.

In the sections which follow, the five main steps included in the present form of the procedures will be described briefly. In order to provide the worker with maximum flexibility at the present stage, in addition to examples of present best practice, the underlying principles for the step will be discussed and also the chief limitations with, wherever possible, suggestions for studies that may result in future improvements in the methods.

1. *General Aims*

A basic condition necessary for any work on the formulation of a functional description of an activity is a fundamental orientation in terms of the general aims of the activity. No planning and no evaluation of specific behaviors are possible without a general statement of objectives. The trend in the scientific field toward operational statements has led a number of writers to try to describe activities or functions in terms of the acts or operations performed, the materials acted on, the situations involved, the results or products, and the relative importance of various acts and results. These analyses have been helpful in emphasizing the need for more specific and detailed descriptions of the requirements of activities. Typically, however, such discussions have failed to emphasize the dominant role of the general aim in formulating a description of successful behavior or adjustment in a particular situation.

In its simplest form, the functional description of an activity specifies

precisely what it is necessary to do and not to do if participation in the activity is to be judged successful or effective. It is clearly impossible to report that a person has been either effective or ineffective in a particular activity by performing a specific act unless we know what he is expected to accomplish. For example, a supervisor's action in releasing a key worker for a half a day to participate in a recreational activity might be evaluated as very effective if the general aim of the foreman was to get along well with the employees under him. On the other hand, this same action might be evaluated as ineffective if the primary general aim is the immediate production of materials or services.

In the case of the usual vocational activities the supervisors can be expected to supply this orientation. In certain other types of activities, such as civic, social, and recreational activities, there frequently is no supervisor. The objectives of participation in the activity must then be determined from the participants themselves. In some instances, these may not be verbalized to a sufficient extent to make it possible to obtain them directly.

Unfortunately, in most situations there is no one general aim which is the correct one. Similarly, there is rarely one person or group of persons who constitute an absolute, authoritative source on the general aim of the activity. In a typical manufacturing organization the foreman, the plant manager, the president, and the stockholders might define the general aim of the workers in a particular section somewhat differently. It is not possible to say that one of these groups knows the correct general aim and the others are wrong. This does not mean that one general aim is as good as another and that it is unim-

portant how we define the purpose of the activity. It does mean that we cannot hope to get a completely objective and acceptable general aim for a specific activity. The principal criterion in formulating procedures for establishing the general aim of the activity should be the proposed use of the functional description of the activity which is being formulated. Unless the general aim used is acceptable to the potential users of the detailed statement of requirements, the whole effort in formulating this statement will have been wasted.

The most useful statements of aims seem to center around some simple phrase or catchword which is slogan-like in character. Such words provide a maximum of communication with only a minimum of possible misinterpretation. Such words as "appreciation," "efficiency," "development," "production," and "service" are likely to be prominent in statements of general aims. For example, the general aim of a teacher in elementary school art classes might be the development of an appreciation of various visual art forms on the part of the students. The general aim of the good citizen might be taken as effective participation in the development and application of the rules and procedures by which individuals and groups are assisted in achieving their various goals.

With the aid of a form of the type shown in Fig. 1, the ideas of a number of well-qualified authorities can be collected. It is expected that in response to the question on the primary purpose of the activity many persons will give a fairly lengthy and detailed statement. The request to summarize is expected to get them to condense this into a brief usable statement. These should be pooled and a trial form of the statement of general aim developed. This state-

ment should be referred either to these authorities or to others to obtain a final statement of the general aim that is acceptable to them. Necessary revisions should be made as indicated by these discussions. Usually considerable effort is required to avoid defeating the purpose of the

OUTLINE FOR INTERVIEW TO
ESTABLISH THE GENERAL
AIM FOR AN ACTIVITY

1. *Introductory statement:* We are making a study of (specify activity). We believe you are especially well qualified to tell us about (specify activity).
2. *Request for general aim:* What would you say is the primary purpose of (specify activity)?
3. *Request for summary:* In a few words, how would you summarize the general aim of (specify activity)?

FIG. 1. SAMPLE FORM FOR USE IN OBTAINING GENERAL AIM

general aim by cluttering up the statement with specific details and qualifying conditions.

In summary, the general aim of an activity should be a brief statement obtained from the authorities in the field which expresses in simple terms those objectives to which most people would agree. Unless a brief, simple statement has been obtained, it will be difficult to get agreement among the authorities. Also it will be much harder to convey a uniform idea to the participants. This latter group will get an over-all impression and this should be as close to the desired general aim as possible.

2. *Plans and Specifications*

To focus attention on those aspects of behavior which are believed to be crucial in formulating a functional description of the activity, precise

instructions must be given to the observers. It is necessary that these instructions be as specific as possible with respect to the standards to be used in evaluation and classification. The group to be studied also needs to be specified.

One practical device for obtaining specific data is to obtain records of "critical incidents" observed by the reporting personnel. Such incidents are defined as extreme behavior, either outstandingly effective or ineffective with respect to attaining the general aims of the activity. The procedure has considerable efficiency because of the use of only the extremes of behavior. It is well known that extreme incidents can be more accurately identified than behavior which is more nearly average in character.

One of the primary aims of scientific techniques is to insure objectivity for the observations being made and reported. Such agreement by independent observers can only be attained if they are all following the same set of rules. It is essential that these rules be clear and specific. In most situations the following specifications will need to be established and made explicit prior to collecting the data:

a. The situations observed. The first necessary specification is a delimitation of the situations to be observed. This specification must include information about the place, the persons, the conditions, and the activities. Such specifications are rather easily defined in many instances. For example, such brief specifications as observations of "the behavior in classrooms of regularly employed teachers in a specified high school while instructing students during class periods," constitute a fairly adequate definition of a situation of this type.

In complex situations it is probably essential not only that the specifications with respect to the situation be relatively complete and specific, but also that practical examples be provided to assist the observer in deciding in an objective fashion whether or not a specific behavior should be observed and recorded.

b. Relevance to the general aim. After the decision has been made that a particular situation is an appropriate one for making observations, the next step is to decide whether or not a specific behavior which is observed is relevant to the general aim of the activity as defined in the section above. For example, if the general aim of the activity was defined as sustained high quality and quantity of production, it might be difficult to decide whether or not to include an action such as encouraging an unusually effective subordinate to get training that would assist him in developing his ability in an avocational or recreational activity not related to his work. In this case, it might be specified that any action which either directly or indirectly could be expected over a long period of time to have a significant effect on the general aim should be included. If it could not be predicted with some confidence whether this effect would be good or bad, it should probably not be considered.

The extent of detail required to obtain objectivity with respect to this type of decision depends to a considerable degree on the background and experiences of the observers with respect to this activity. For example, supervisors with substantial experience in a particular company can be expected to agree on whether or not a particular behavior is relevant to the attainment of the general aim. On the other hand, if outside observers were to be used, it would probably be necessary to specify in considerable detail the activities that can be expected to have an effect on the general aim.

c. Extent of effect on the general aim. The remaining decision that the observer must make is how important an effect the observed incident has on the general aim. It is necessary to specify two points on the scale of importance: (a) a level of positive contributions to the general aim in specific terms, preferably including a concrete example, and (b) the corresponding level of negative effect on the general aim expressed in similar terms.

A definition which has been found useful is that an incident is critical if it makes a "significant" contribution, either positively or negatively, to the general aim of the activity. The definition of "significant" will depend on the nature of the activity. If the general aim of the activity is in terms of production, a significant contribution might be one which caused, or might have caused, an appreciable change in the daily production of the department either in the form of an increase or a decrease.

In certain specific situations, it might be desirable and possible to set up a quantitative criterion such as saving or wasting 15 minutes of an average worker's production. In some situations, a definition of significance might be set up in terms of dollars saved or lost both directly and indirectly.

Actions which influence the attitudes of others are more difficult to evaluate objectively. Perhaps the best we might be able to do is to state it in terms of a probability estimate. For example, one such criterion might be that the minimum critical level would be an action that would have an influence such that at least one person in ten might change his vote on an issue of importance to the company.

d. Persons to make the observations. One additional set of specifications refers to the selection and training of the observers who are to make and report the judgments outlined in the steps above.

Wherever possible, the observers should be selected on the basis of their familiarity with the activity. Special consideration should be given to observers who have made numerous observations on persons engaged in the activity. Thus, for most jobs, by far the best observers are supervisors whose responsibility it is to see that the particular job being studied is done. However, in some cases very useful observations can be contributed by consumers of the products and services of the activity. For example, for a study of effective sales activities, the customers may have valuable data to contribute. For a study of effective parental activity, the children may be able to make valuable contributions.

In addition to careful selection of the persons to make observations, attention should be given to their training. Minimal training should include a review of the nature of the general aim of the activity and a study of the specifications and definitions for the judgments they will be required to make. Where the situation is complex or the observers are not thoroughly familiar with the activity, supervised practice in applying these definitions should be provided. This can be done by preparing descriptions of observations and asking the observers to make judgments about these materials. Their judgments can be immediately confirmed or corrected during such supervised practice periods.

In Fig. 2 is shown a form for use in developing specifications regarding observations. The use of this form in making plans for the collection of critical incidents or other types of observational data should aid in objectifying these specifications.

3. *Collecting the Data*

If proper plans and specifications are developed, the data collection phase is greatly simplified. A necessary condition for this phase is that the behaviors or results observed be evaluated, classified, and recorded while the facts are still fresh in the mind of the observer. It would be desirable for these operations to be

---

Specifications Regarding Observations

1. Persons to make the observations.
   a. Knowledge concerning the activity.
   b. Relation to those observed.
   c. Training requirements.
2. Groups to be observed.
   a. General description.
   b. Location.
   c. Persons.
   d. Times.
   e. Conditions.
3. Behaviors to be observed.
   a. General type of activity.
   b. Specific behaviors.
   c. Criteria of relevance to general aim.
   d. Criteria of importance to general aim (critical points)

FIG. 2. FORM FOR DEVELOPING SPECIFICATIONS REGARDING OBSERVATIONS

---

performed at the time of observation so that all requisite facts could be determined and checked. Memory is improved if it is known in advance that the behavior to be observed is to be remembered. It is greatly improved if the specific aspects of what is to be observed are defined and if the operations to be performed with respect to evaluation and classification are clearly specified.

The critical incident technique is frequently used to collect data on observations previously made which are reported from memory. This is usually satisfactory when the in-

cidents reported are fairly recent and the observers were motivated to make detailed observations and evaluations at the time the incident occurred.

The importance of obtaining recent incidents to insure that the incidents are representative of actual happenings was demonstrated in the study on air route traffic controllers by Nagay (48) reported above. However, as also discussed in that study, in some situations adequate coverage cannot be obtained if only very recent incidents are included.

Evidence regarding the accuracy of reporting is usually contained in the incidents themselves. If full and precise details are given, it can usually be assumed that this information is accurate. Vague reports suggest that the incident is not well remembered and that some of the data may be incorrect. In several situations there has been an opportunity to compare the types of incidents reported under two conditions (a) from memory and without a list of the types of incidents anticipated, and (b) those reported when daily observations were being made in a routine work situation, and the evaluations and classifications were made and recorded on a prepared check list within 24 hours of the time of observation. The results of one such comparison were discussed briefly above in connection with the American Institute for Research study of factory employees.

During the observational period a negligible number of incidents were reported by the foremen as not fitting into the general headings included on the list. Although the proportions of incidents for the various items on the list are not identical, they are reasonably close for most of the items. Items on such matters as meeting production requirements and accepting changes in jobs are higher in terms of the recorded than the recalled incidents. The fact that items such as wasting time and assisting on problems are lower for the recalled incidents suggests that part of this discrepancy lies in the interpretations of the category definitions. The classifying of recorded incidents was done by the foremen, while the classification of the recalled incidents was done by the research workers. In fairness, it should also be noted that the definitions used by the research workers were rewritten before they were incorporated in the foremen's manuals.

On the whole, it seems reasonable to assume that, if suitable precautions are taken, recalled incidents can be relied on to provide adequate data for a fairly satisfactory first approximation to a statement of the requirements of the activity. Direct observations are to be preferred, but the efficiency, immediacy, and minimum demands on cooperating personnel which are achieved by using recalled incident data frequently make their use the more practical procedure.

Another practical problem in collecting the data for describing an activity refers to the problem of how it should be obtained from the observers. This applies especially to the problem of collecting recalled data in the form of critical incidents. Four procedures have been used and will be discussed briefly below:

a. Interviews. The use of trained personnel to explain to observers precisely what data are desired and to record the incidents, making sure that all necessary details are supplied, is probably the most satisfactory data collection procedure. This type of interview is somewhat different from other sorts of interviews and a brief summary of the principal factors involved will be given.

(i) *Sponsorship of the study.* If a stranger to

the observers is collecting the data, it is ordinarily desirable to indicate on what authority the interview is being held. This part should be as brief as possible to avoid any use of time for a prolonged discussion of a topic irrelevant to the purpose of the interview. In many instances all that needs to be said is that someone known and respected by the observer has suggested the interview.

(ii) *Purpose of the study*. This should also be brief and ordinarily would merely involve a statement that a study was being made to describe the requirements of the activity. This would usually be cast in some such informal form as, "We wish to find out what makes a good citizen," or, "We are trying to learn in detail just what successful work as a nurse includes." In cases where there is some hesitation about cooperating or a little more explanation seems desirable, a statement can be added concerning the value and probable uses of the results. This frequently takes the form of improving selection and training procedures. In some instances, it would involve improving the results of the activity. For example, the interviewer might say, "In order to get better sales clerks we need to know just what they do that makes them especially effective or ineffective," or, "If parents are to be more effective, we need to be able to tell them the things they do that are effective and ineffective."

(iii) *The group being interviewed*. If there is any likelihood of a person feeling, "But, why ask me?" it is desirable to forestall this by pointing out that he is a member of a group which is in an unusually good position to observe and report on this activity. The special qualifications of members of this group as observers can be mentioned briefly, as, "Supervisors such as yourself are constantly observing and evaluating the work of switchboard girls," or, "Students are in an unusually good position to observe the effectiveness of their teachers in a number of ways."

(iv) *The anonymity of the data*. Especially for the collection of information about ineffective behavior, one of the principal problems is to convince the observer that his report cannot harm the person reported on in any way. Usually he also needs to be convinced that the person reported on will never know that he has reported the incident. Assurances are not nearly so effective in this situation as actual descriptions of techniques to be used in handling the data, which enable the observer to judge for himself how well the anonymity of the data will be guarded. Under no circumstances should the confidences of the reporters be violated in any way. The use of sealed envelopes, avoidance of identifying information, the mailing of data immediately to a distant point for analysis, and similar techniques are helpful in establishing the good faith of the interviewer in taking all possible precautions to safeguard the incidents reported.

(v) *The question*. The most crucial aspect of the data collection procedure is the questions asked the observers. Many studies have shown that a slight change in wording may produce a substantial change in the incidents reported. For example, in one study the last part of one of the specific questions asked was, "Tell just how this employee behaved which caused a noticeable decrease in production." This question resulted in almost all incidents reported having to do with personality and attitude behaviors. This part of the question was changed to, "Tell just what this employee did which caused a noticeable decrease in production." This second question produced a much broader range of incidents. To the question writer "how he behaved" and "what he did" seemed like about the same thing. To the foremen who were reporting incidents "how he behaved" sounded as if personality and attitudes were being studied. The subtle biases involved in the wording of questions are not always so easily found. Questions should always be tried out with a small group of typical observers before being put into general use in a study.

The question should usually refer briefly to the general aim of the activity. This aim might be discussed more fully in a preliminary sentence. It should usually state that an incident, actual behavior, or what the person did is desired. It should briefly specify the type of behavior which is relevant and the level of importance which it must reach to be reported. It should also tie down the selection of the incidents to be reported by the observer in some way, such as asking for the most recent observation, in order to prevent the giving of only the more dramatic or vivid incidents, or some other selected group, such as those which fit the observer's stereotypes.

An effective procedure for insuring that the interpretation of the persons being interviewed is close to that intended is to request a sample of persons typical of those to be interviewed to state in their own words what they understand they have been asked to do. These persons should be selected so as to represent all types who will be interviewed. From a study of their interpretations, necessary revisions can be made to insure that all interviewees will be in agreement as to the nature of the incidents they are to provide.

(vi) *The conversation*. The interviewer

should avoid asking leading questions after the main question has been stated. His remarks should be neutral and permissive and should show that he accepts the observer as the expert. By indicating that he understands what is being said and permitting the observer to do most of the talking, the interviewer can usually get unbiased incidents. If the question does not seem to be understood, it can be repeated with some reference to clarifying just what is meant by it. If the observer has given what seems like only part of the story, he should be encouraged by restating the essence of his remarks. This usually tends to encourage him to continue and may result in his bringing out many relevant details that the interviewer did not know the situation well enough to ask for. In some cases, it is desirable to have the interviews recorded electrically and transcribed. This increases the work load substantially, and trained interviewers can usually get satisfactory reports at the time or by editing their notes shortly after the interview.

Usually the interviewer should apply certain criteria to the incidents while they are being collected. Some of the more important criteria are: (a) is the actual behavior reported; (b) was it observed by the reporter; (c) were all relevant factors in the situation given; (d) has the observer made a definite judgment regarding the criticalness of the behavior; (e) has the observer made it clear just why he believes the behavior was critical.

In Fig. 3 is shown a sample of the type of form used by interviewers to collect critical incidents. Of course the form must be adapted to the needs of the specific situation.

b. *Group interviews.* Because of the cost in time and personnel of the individual interview, a group interview technique has been developed. This retains the advantages of the individual interview in regard to the personal contact, explanation, and availability of the interviewer to answer questions. To some extent it also provides for a check on the data supplied by the interviewees. Its other advantages are that the language of the actual observer is precisely reproduced and the time for editing the interviews is virtually eliminated.

The method consists of having the interviewer give his introductory remarks to a group very much as he would do in an individual interview. There is an opportunity for questions and clarification. Then each person is asked to write incidents in answer to specific questions contained on a specially prepared form. The size of the group which can be handled effectively will vary with the situation. If the group is fairly small, it is usually possi-

"Think of the last time you saw one of your subordinates do something that was very helpful to your group in meeting their production schedule." (Pause till he indicates he has such an incident in mind.) "Did his action result in increase in production of as much as one per cent for that day?—or some similar period?"

(If the answer is "no," say) "I wonder if you could think of the last time that someone did something that did have this much of an effect in increasing production." (When he indicates he has such a situation in mind, say) "What were the general circumstances leading

up to this incident?" _____

_____

"Tell me exactly what this person did that was so helpful at that time."

_____

"Why was this so helpful in getting

your group's job done?" _____

_____

"When did this incident happen?"

_____

"What was this person's job?" _____

_____

"How long has he been on this job?"

_____

"How old is he?" _____

Fig. 3. Sample of a Form for Use by an Interviewer in Collecting Effective Critical Incidents

ble for the interviewer to read the responses of each member of the group to the first question and make sure that he understands what is wanted. There seems to be a certain amount of social facilitation, and the results in most situations have been excellent. In the report of the first use of this procedure by Wagner (65), the amount of interviewer time required per usable incident was 4.3 minutes for the group interview procedure as compared with 15.7 minutes for individual interviews. The

quantity of ... incidents, obtained from various ... States ... appeared to be about the same for the two situations.

c. Questionnaires. If the group becomes large, the group interview procedure is more in the nature of a questionnaire procedure. There are, of course, all types of combinations of procedures that can be used. The one that is most different from those discussed is the mailed questionnaire. In situations where the observers are motivated to read the instructions carefully and answer conscientiously, this technique seems to give results which are not essentially different from those obtained by the interview method. Except for the addition of introductory remarks, the forms used in collecting critical incidents by means of mailed questionnaires are about the same as those used in group interviews.

d. Record forms. One other procedure for collecting data is by means of written records. There are two varieties of recording: one is to record details of incidents as they happen. This situation is very similar to that described in connection with obtaining incidents by interviews above, except that the observation and giving of incidents are delayed following the introductory remarks and the presentation of the questions until an incident is observed to happen.

A variation of this procedure is to record such incidents on forms which describe most of the possible types of incidents by placing a check or tally in the appropriate place.

As additional information becomes available on the nature of the components which make up activities, observers may thus collect data more efficiently by using forms for recording and classifying observations. In the meantime, because of the inadequacy of the information currently available regarding these components, it seems desirable to ask observers to report their observations in greater detail and have the classification done by specially trained personnel.

Size of sample. A general problem which overlaps the phases of collecting the incidents and analyzing the data relates to the number of incidents required. There does not appear to be a simple answer to this question. If the activity or job being defined is relatively simple, it may be satisfactory to collect only 50 or 100 incidents. On the other hand, some types of complex activity appear to require several thousand incidents for an adequate statement of requirements.

The most useful procedure for determining whether or not additional incidents are needed is to keep a running count on the number of new critical behaviors added to the classifica-

tion system with each additional 100 incidents. In the ... progress ... it seemed that adequate ... has been obtained when the addition of 100 critical incidents to the sample adds only two or three critical behaviors. For jobs of a supervisory nature, it appears that between 2,000 and 4,000 critical incidents are required to establish a comprehensive statement of requirements that includes nearly all of the different types of critical behaviors. For semiskilled and skilled jobs between 1,000 and 2,000 incidents seem to be adequate to cover the critical behaviors.

Coverage of all or nearly all of the various critical behaviors is not the only criterion as to whether or not a sufficient number of critical incidents has been collected. If a relatively precise definition of each critical behavior category is required, it may be necessary to get at least three or four examples of each critical behavior. Similarly, if the critical incidents are to be used as a basis for developing selection tests, training materials, and proficiency measures, more incidents may be required to provide a sufficient supply of usable ideas for the development of these materials.

In summary, although there is no simple formula for determining the number of critical incidents that will be required, this is a very important consideration in the plan of the study; checks should be made both on the first hundred or so incidents and again after approximately half of the number of incidents believed to be required have been obtained in order to make it possible to revise the preliminary estimates, if necessary, with a minimum loss in effort and time.

4. *Analyzing the Data*

The collection of a large sample of incidents that fulfill the various conditions outlined above provides a functional description of the activity in terms of specific behaviors. If the sample is representative, the judges well qualified, the types of judgments appropriate and well defined, and the procedures for observing and reporting such that incidents are reported accurately, the stated requirements can be expected to be comprehensive, detailed, and valid in this form. There is only one reason for going further and that is practical utility. The purpose of the data analysis

stage is to summarize and describe the data in an efficient manner so that it can be effectively used for many practical purposes.

In the discussion which follows, it should be kept in mind that the process of description has been completed. The specific procedures to be discussed are not concerned with improving on the comprehensiveness, specificity of detail, or validity of the statement of the requirements of the activity. Rather, they are concerned with making it easier to report these requirements, to draw inferences from them, and to compare the activity with other activities.

The aim is to increase the usefulness of the data while sacrificing as little as possible of their comprehensiveness, specificity, and validity. It appears that there are three primary problems involved: (a) the selection of the general frame of reference that will be most useful for describing the incidents; (b) the inductive development of a set of major area and subarea headings; and (c) the selection of one or more levels along the specificity-generality continuum to use in reporting the requirements. Each of these problems will be discussed below:

a. *Frame of reference.* There are countless ways in which a given set of incidents can be classified. In selecting the general nature of the classification, the principal consideration should usually be that of the uses to be made of the data. The preferred categories will be those believed to be most valuable in using the statement of requirements. Other considerations are ease and accuracy of classifying the data, relation to previously developed definitions or classification systems, and considerations of interpretation and reporting, which will be discussed in a later section.

For job activities, the choice of a frame of reference is usually dominated by considerations of whether the principal use of the requirements will be in relation to selection, training, measurement of proficiency, or the

development of procedures for evaluat⟨…⟩ the-job effectiveness. For selection ⟨…⟩ the most appropriate classification ⟨…⟩ psychological one. The main headings have ⟨…⟩ do with types of psychological traits that ⟨…⟩ utilized in the selection process. For training uses, the best classification system follows a set of headings that is easily related to training courses or broad training aims. For proficiency measurement, the headings tend to be similar to those for training except that there is less attention to possible course organization and aims and greater attention to the components of the job as it is actually performed. For the development of procedures for evaluating on-the-job effectiveness to establish a criterion of success, the classification system is necessarily directed at presenting the on-the-job behaviors under headings that represent either well-marked phases of the job or provide a simple framework for classifying on-the-job activities that is either familiar to or easily learned by supervisors.

Similarly, in nonvocational activities the frame of reference depends on the uses planned for the findings. For example, if a study is being made to define immaturity reactions in military personnel, the frame of reference would depend somewhat on whether the functional description is to be used primarily to identify personnel showing this type of maladjustment or whether the principal use will be to try to prepare specifications for types of situations in which immaturity reactions would not lead to serious difficulties.

b. *Category formulation.* The induction of categories from the basic data in the form of incidents is a task requiring insight, experience, and judgment. Unfortunately, this procedure is, in the present stage of psychological knowledge, more subjective than objective. No simple rules are available, and the quality and usability of the final product are largely dependent on the skill and sophistication of the formulator. One rule is to submit the tentative categories to others for review. Although there is no guarantee that results agreed on by several workers will be more useful than those obtained from a single worker, the confirmation of judgments by a number of persons is usually reassuring. The usual procedure is to sort a relatively small sample of incidents into piles that are related to the frame of reference selected. After these tentative categories have been established, brief definitions of them are made, and additional incidents are classified into them. During this process, needs for redefinition and for the development of new categories are noted. The

tentative categories are modified as indicated and the process continued until all the incidents have been classified.

The larger categories are subdivided into smaller groups and the incidents that describe very nearly the same type of behavior are placed together. The definitions for all the categories and for their headings should then be re-examined in terms of the actual incidents classified under each.

c. General behaviors. The last step is to determine the most appropriate level of specificity-generality to use in reporting the data. This is the problem of weighing the advantages of the specificity achieved in specific incidents against the simplicity of a relatively small number of headings. The level chosen might be only a dozen very general behaviors or it might be several hundred rather specific behaviors. Practical considerations in the immediate situation usually determine the optimal level of generality to be used.

Several considerations should be kept in mind in establishing headings for major areas and in stating critical requirements at the selected level of generality. These are listed below:

i. The headings and requirements should indicate a clear-cut and logical organization. They should have a discernible and easily remembered structure.

(ii) The titles should convey meanings in themselves without the necessity of detailed definition, explanation, or differentiation. This does not mean that they should not be defined and explained. It does mean that these titles, without the detailed explanation, should still be meaningful to the reader.

(iii) The list of statements should be homogeneous; i.e., the headings for either areas or requirements should be parallel in content and structure. Headings for major areas should be neutral, not defining either unsatisfactory or outstanding behaviors. Critical requirements should ordinarily be stated in positive terms.

(iv) The headings of a given type should all be of the same general magnitude or level of importance. Known biases in the data causing one area or one requirement to have a disproportionate number of incidents should not be reflected in the headings.

(v) The headings used for classification and reporting of the data should be such that findings in terms of them will be easily applied and maximally useful.

(vi) The list of headings should be comprehensive and cover all incidents having significant frequencies.

5. *Interpreting and Reporting*

It is never possible in practice to obtain an ideal solution for each of the practical problems involved in obtaining a functional description of an activity. Therefore, the statement of requirements as obtained needs interpretation if it is to be used properly. In many cases, the real errors are made not in the collection and analysis of the data but in the failure to interpret them properly. Each of the four preceding steps, (a) the determination of the general aim, (b) the specification of observers, groups to be observed, and observations to be made, (c) the data collection, and (d) the data analysis, must be studied to see what biases have been introduced by the procedures adopted. If there is a division of opinion as to the general aim and one of the competing aims is selected, this should be made very clear in the report. If the groups on whom the observations are made are not representative of the relevant groups involved, they must be described as precisely as possible. The aim of the study is usually not a functional description of the activity as carried on by this sample but rather a statement relating to all groups of this type. In order to avoid faulty inferences and generalizations, the limitations imposed by the group must be brought into clear focus. Similarly, the nature of judgments made in collecting and analyzing the data must be carefully reviewed.

While the limitations need to be clearly reported, the value of the results should also be emphasized. Too often the research worker shirks his responsibility for rendering a judgment concerning the degree of credibility which should be attached to his findings. It is a difficult task,

but if the results are to be used, someone will have to make such a judgment, and the original investigator is best prepared to make the necessary evaluations either for the general case or for certain typical specific examples.

USES OF THE CRITICAL INCIDENT TECHNIQUE

The variety of situations in which the collection of critical incidents will prove of value has only been partially explored. In the approximately eight years since the writer and his colleagues began a systematic formulation of principles and procedures to be followed in collecting this type of data, a fairly large number of applications has been made. The applications will be discussed under the following nine headings: (a) Measures of typical performance (criteria); (b) measures of proficiency (standard samples); (c) training; (d) selection and classification; (e) job design and purification; (f) operating procedures; (g) equipment design; (h) motivation and leadership (attitudes); (i) counseling and psychotherapy.

Space is not available here to describe these various applications in detail. However, a brief description of the types of application that have been made, along with brief illustrative examples and references, will be presented. Some of the studies involve several of the types of applications to be discussed. The presentation is not intended to be complete, but rather to give the reader interested in further study some orientation and guidance.

Measures of typical performance (criteria). The simplest and most natural application of a systematically collected set of critical incidents is in terms of the preparation of a statement of critical requirements

and a check list or some similar type of procedure for evaluating the typical performance of persons engaged in this activity. If an observational check list that includes all of the important behaviors for the activity is available, the performance of the individual can be objectively evaluated and recorded by merely making a single tally mark for each observation. Such records provide the essential basis for criterion data which are sufficiently detailed and specific for special purposes but at the same time can be combined into a single over-all evaluation when this is desirable. Such a procedure was first suggested and tried out in connection with developmental studies of the American Institute for Research. These included: Preston's study of officers for the United States Air Force (52); Nagay's study on air route traffic controllers for the Civil Aeronautics Administration (49); and M. H. Weislogel's study on research personnel for the Office of Naval Research (69). Another American Institute for Research study was reported by R. B. Miller and the present author (21). This was a performance record form for hourly wage employees developed in cooperation with personnel of the Delco-Remy Division of the General Motors Corporation, the Employment Practices Division of that corporation, and the Industrial Relations Center of the University of Chicago. The same authors have developed similar performance records for salaried employees, and foremen and supervisors (22, 23). The principles and procedures underlying this type of evaluation of performance have been published elsewhere (14, 15, 17).

A number of important contributions to the development of functional descriptions and standards of

performance have been made by other groups using the critical incident technique. One of the most notable of these is the development by Hobbs *et al.* (3, 31), of *Ethical Standards of Psychologists.* More than 1,000 critical incidents involving ethical problems of psychologists were contributed by the members of the American Psychological Association. It is believed that this represents the first attempt to use empirical methods to establish ethical standards. Because of the importance of this study, and the generality of some of the problems involved, certain of the conclusions reported by the Committee on Ethical Standards for Psychology in their introductory statement will be quoted here.

First, it is clear that psychologists believe that ethics are important; over two thousand psychologists were sufficiently concerned with the ethical obligations of their profession to contribute substantially to the formulation of these ethical standards. Second, psychologists believe that the ethics of a profession cannot be prescribed by a committee; ethical standards must emerge from the day-by-day value commitments made by psychologists in the practice of their profession. Third, psychologists share a conviction that the problems of men, even those involving values, can be studied objectively; this document summarizes the results of an effort to apply some of the techniques of social science to the study of ethical behavior of psychologists. Fourth, psychologists are aware that a good code of ethics must be more than a description of the current status of ethics in the profession; a code must embody the ethical aspirations of psychologists and encourage changes in behavior, bringing performance ever closer to aspiration. Fifth, psychologists appreciate that process is often more important than product in influencing human behavior; the four years of widely-shared work in developing this code are counted on to be more influential in changing ethical practices of psychologists than will be the publication of this product of their work. Finally, psychologists recognize that the process of studying ethical standards must be a continuing one; occasional publications such as this statement mark no point of conclusion in the ongoing process of defining ethical standards

— they are a means of sharing the more essential discipline of examining professional experience, forming hypotheses about professional conduct, and testing these hypotheses by reference to the welfare of the people affected by them (3, p. v).

In addition to the study by Smit mentioned in a previous section (58), several other studies on the use of the critical incident procedures as a basis for evaluating teaching effectiveness have been reported. One of these was a study conducted under the joint sponsorship of the Educational Research Corporation and the Harvard University Graduate School of Education with funds provided by the New England School Development Council and the George F. Milton Fund. This was an exploratory study of teacher competence reported by Domas (6). Approximately 1,000 critical incidents were collected from teachers, principals, and other supervisors. Although this was an exploratory study, it was felt that it made an important contribution to the general problem of relating salary to teacher competence.

The second of these studies was conducted as part of the teacher characteristics study sponsored by the American Council on Education and subsidized by the Grant Foundation. This study is reported by Jensen (32). Teachers, administrators, and teachers in training in the Los Angeles area contributed more than 1,500 critical incidents of teacher behavior. The incidents were classified under personal, professional, and social qualities. The category formulation indicated that there were about 20 distinct critical requirements. These were recommended as a basis for teacher evaluation and as an aid to the in-service growth of teachers.

Another study was that of Smith and Staudohar (59), which deter-

mined the critical requirements for basic training of tactical instructors in the United States Air Force. From 130 training supervisors, 555 tactical instructors, and 3,082 basic trainees, a total of 6,615 usable incidents were obtained. The authors comment that:

> The training supervisors report a predominance of ineffective incidents in the major areas of: Sets a good example and maintains effective personal relations. The tactical instructors report more ineffective incidents in the area of Makes his expectations clear. Basic trainees show a predominance of ineffective incidents in three areas: Sets a good example, Considers trainee's needs, and Maintains effective personal relations (59, p. 5).

Another study on the evaluation of instructor effectiveness was carried out by Konigsburg (33). This study involved the development of an instructor check list for college instructors based on the critical incident technique and a comparison of techniques for recording observations. Its principal findings were the very low correlation coefficients between the total scores from the Purdue Rating Scale for Instruction and the instructor check list. When these two instruments were each given to half the class on the same day, the average correlation coefficient was found to be .29. The other principal finding is that the planned performances of a total of 46 predetermined behaviors were better reflected by the results obtained on the instructor check list than by the results on the Purdue Rating Scale.

A somewhat related study has been reported by Barnhart (4). This study collected a large number of critical incidents for the purpose of establishing the critical requirements for school board membership. The author applied his findings to the problem of evaluating the effectiveness of school board members.

Another type of application of the critical incident technique to the development of bases for evaluating behavior is the previously mentioned study of Eilbert (7). His list of 51 types of immature reaction based on a collection of several hundred critical incidents describing manifestations of emotional immaturity is believed to provide a useful guide to further investigation and appraisal of persons with behavior problems. It is believed that the results of this study provide substantial encouragement to the application of the critical incident technique to similar problems in the field of clinical diagnosis and evaluation.

Measures of proficiency (standard samples). A closely related use of critical incidents is to provide a basis for evaluating the performance of persons by use of standard samples of behavior involving important aspects of the activity. Such evaluations are called proficiency measures and are differentiated from the evaluation of typical performance on the job primarily on the basis that a test situation rather than a real job situation is used. Measures of this sort are especially useful at the end of training courses as checks on the maintenance of proficiency, and when the tasks assigned to participants vary a great deal in difficulty or are not directly observed by the supervisors.

One of the first applications of critical incidents to the development of proficiency measures was Gordon's study on the development of a standard flight check for the airline transport rating (28, 29). This study was done by the American Institute for Research under the sponsorship of the National Research Council Committee on Aviation Psychology with funds provided by the Civil Aeronautics Administration. In this

study data from analyses of airline accidents were combined with critical incidents reported by airline pilots to provide the basis for developing an objective measure of pilot proficiency. The flight check consisted of the presentation of situations providing uniformly standardized opportunities to perform the critical aspects of the airline pilot's job as indicated from the study of the accidents and critical incidents reported. The new check was found to yield 88 per cent agreement on the decision to pass or fail a particular pilot when examined on flights on successive days by different check pilots. The previous flight check when used on the same flights gave only 63 per cent agreement, which was little better than chance under the conditions of the study.

Similar studies on the development of flight checks at the American Institute for Research have been carried out by Marley (36, 37), G. S. Miller (39), and Ericksen (9). These studies, sponsored by the United States Air Force and the Civil Aeronautics Administration, were concerned respectively with objective flight checks for B-29 bombing crew members, B-36 bombing crew members, and private pilots flying light civilian aircraft. Ericksen also developed a light plane proficiency check to predict military flying success (10) on a similar project sponsored by the United States Air Force Human Resources Research Center.

A similar set of proficiency measures was developed by Krumm for Air Force pilot instructors (34, 35), also under the sponsorship of the Human Resources Research Center. These measures were based on more than 4,000 critical incidents collected from student pilots, flight instructors, and supervisors. The critical incidents were classified under three main headings: (a) proficiency as a pilot; (b) proficiency as a teacher; and (c) proficiency in maintaining effective personnel relations. The proficiency measures developed in connection with this study included paper-and-pencil tests presenting critical situations and requiring the instructor to select one of several proposed solutions.

Another development of this type carried on at the American Institute for Research was the construction of tests for evaluating research proficiency in physics and chemistry for the Office of Naval Research by M. H. Weislogel (71). This study was based on the critical incidents for research personnel (20) discussed in a previous section. The items for these proficiency measures were based on detailed rationales. The items described a practical research situation in considerable detail and outlined five specific choices concerning such matters as the best thing to do next, suggestions for improving the procedure as reported, etc. The critical behaviors tested in the items were taken directly from the critical incidents. The method of developing tests through the use of comprehensive rationales has been discussed generally in another paper (16).

Three studies have been reported by the American Institute for Research in which critical incidents were used as a basis for developing situational performance tests for measuring certain aspects of the proficiency of military personnel. These included the study of Sivy and Lange on the development of an objective form of the Leaders Reaction Test for the Personnel Research Branch, Department of the Army (57). This test included 20 situational problems based on the critical requirements of

the noncommissioned combat infantry leader as determined on the basis of critical incidents collected in military maneuvers and during combat operations at the front in Korea. A second proficiency measure of a somewhat similar sort was developed for other types of personnel by R. L. Weislogel (73). The third study of this type was carried out by Suttell (61) for the Human Resources Research Center. This study was based on critical incidents collected in previous studies of the American Institute for Research and reported the development and preliminary evaluation of the Officer Situations Test. This test was designed to measure nonintellectual aspects of officer performance through the use of 16 situational problems requiring about six hours of testing time.

Because of the great difficulty in obtaining valid and reliable measures of typical performance, accurate measures of proficiency are essential for many types of activities. It is apparent that a comprehensive set of critical incidents can be of great value in constructing such measures.

Training. Many of the applications of the critical incident technique to training problems have been carried out for the military in special situations so that the reports are classified security information. In addition to work by Preston, Glaser, and R. L. Weislogel, R. B. Miller and Folley have utilized critical incidents in establishing training requirements for specific types of maintenance mechanics (47) in a study for the Human Resources Research Center.

Similarly, Ronan has used critical incidents as a basis for developing a program of training for emergency procedures in multi-engine aircraft (54) in a study for the United States

Air Force Human Factors Operations Research Laboratory. On the basis of several thousand incidents reported by aircrew personnel regarding emergencies, three evaluation devices were prepared. These involved a conventional type multiple-choice test; a special multiple-choice test designed to measure the individual's information concerning the important cues in the emergency situation, the appropriate actions to be taken, and the basic troubles or causes of the emergency; and a "flight check" to be used in evaluating the performance of aircrew members in electronic flight simulators.

The obvious relevance of the behaviors involved in critical incidents and the specific details included make such incidents an ideal basis for developing training programs and training materials.

A recent study by Collins (5) uses critical incidents as a basis for evaluating the effectiveness of a training program. The types of incidents reported by mothers after a two-week training course were significantly different from those reported at the beginning of the program in a number of aspects relevant to the objectives of the program. The critical incidents appeared to provide a much more sensitive basis for revealing changes than other procedures used.

Selection and classification. Until recently, the customary approach of the research psychologist to the development of tests for selection and classification purposes has been as follows: A very brief period was given to study of the job. Following this, a wide variety of selection procedures was administered to a group of applicants or employees, and follow-up data were gathered. Since the research psychologist had little confidence in the accuracy of his analysis

a the psychological elements re-
quired by the job there was a tend-
ency to try everything that was
possible and seemed even remotely
related to the tasks involved. This
has been called the "shotgun ap-
proach." It was hoped that with a
wide scatter at least a few of the tests
would pay off. The critical incident
technique has lent substantial sup-
port to the more thorough study of
the job prior to initiating testing pro-
cedures. There is increasing feeling
at the present time that a much
larger percentage of the investiga-
tor's time should be spent on deter-
mining the critical requirements of
the job, so that the psychologist will
have sufficient confidence in his
tentative conclusions as to the nature
of the important selection procedures
to permit their use on a tentative
basis prior to the collection of em-
pirical follow-up data. This is espe-
cially important in those situations
where the follow-up requires a very
long period of time or where the
number of cases that can be followed
up is so small that definitive findings
cannot be anticipated.

One of the most important re-
quirements for developing a system
of job analysis that will facilitate a
relatively accurate identification of
the important job elements for a
specific task is to establish a clear and
specific set of definitions for these job
elements in behavioral terms. The
American Institute for Research has
carried out a series of projects on this
problem. The first of these was a
study undertaken by Wagner under
the sponsorship of the United States
Air Force School of Aviation Medi-
cine to define the requirements of air-
crew jobs in terms of specific job
elements (67, 68). Several thousand
critical incidents were gathered from
aircrew members, and these were

classified into 24 job elements. These
job elements were objectively formu-
lated from the critical incidents and
were grouped under the four area
headings of (a) learning and thinking;
(b) observation and visualization;
(c) sensory-motor coordination and
(d) motives, temperament, and lead-
ership.

The development of more than 100
proficiency tests to measure each of
the various critical behaviors in-
cluded in the 24 tentatively proposed
job elements was reported by Hahn
(30) for the School of Aviation Medi-
cine. These tests were administered
to a group of approximately 500 high
school senior boys, and the intercor-
relations were used to reformulate
the tentative job elements. In a
study just completed by Taylor (62)
for the Human Resources Research
Center, the results of applying an
analytical procedure developed by
Horst to study the interrelationships
involved are reported. This analysis
led to the formulation of a new set of
20 job elements for each of which a
selection test has been developed.
These tests have been administered
to several hundred aviation cadets
and follow-up data on their success in
aircrew training should be available
soon.

A similar project based on critical
incidents collected from various ci-
vilian jobs has been reported by the
present author (2, 18, 19). The
Flanagan Aptitude Classification
Test Series, published in 1953, pro-
vides aptitude measures for 14 criti-
cal job elements. The Applicant
Inventory, also published in 1953,
measures attitudes predictive of job
adjustment for hourly wage em-
ployees.

An effort to adapt the critical in-
cident technique to the problem of
developing civil service examinations

is reported by Wager and Sharon (64). In an exploratory study they collected about 100 incidents regarding on-the-job behaviors of maintenance technicians. These incidents were used as a basis for determining job requirements in terms of behavior, and test items were developed for use in selecting applicants who could be expected to meet these requirements.

Another study that used critical incidents as a basis for developing tests to predict performance was carried out by O'Donnell (51). His test, designed to predict success in dentistry, was based on critical incidents collected by Wagner. The test includes items designed to predict, in part, the following three general areas: (a) demonstrating technical proficiency; (b) handling patient relationships; and (c) accepting professional responsibility. A follow-up study indicated moderate validity for these materials.

One of the few studies known to the author in which the critical incident technique was used in a project carried on outside the United States is Emons' doctor's dissertation (8). This study, carried out at the University of Liége, investigated the aptitudes of effective sales personnel in a large department store. A group of 40 supervisors provided 228 critical incidents. Nine categories were formulated from this group of incidents and recommendations made for an aptitude test to improve current selection procedures.

5. *Job design and purification.* Inadequate attention has been given to the scientific design of jobs to promote over-all efficiency. Where a team has several different types of tasks to perform, it is frequently possible to design each of the team member's jobs so that only a few of the

several tasks are involved. If the jobs have been studied by use of the critical incident technique, it may be possible to select and train each team member for only two or three of the critical job elements. This tends to maximize the effectiveness of performance with respect to each of the various types of tasks. Although such procedures have nearly always been informally used in planning the work of teams, the critical incident technique facilitates the collection of the data essential to this type of job purification.

Some preliminary work on this problem has been carried out at the American Institute for Research. Recommendations resulting from these studies for reducing the number of job elements required in certain common maintenance jobs are expected to lead to a saving of millions of dollars in training costs as well as to improving the effectiveness of job performance.

Operating procedures. Another application of critical incidents which has not been adequately exploited is the study of operating procedures. Detailed factual data on successes and failures that can be systematically analyzed are of great importance in improving the effectiveness and efficiency of operations. Such information can be efficiently collected by means of the critical incident technique.

Examples of such studies are provided by a series of three projects carried out by the American Institute for Research under the sponsorship of the United States Air Force School of Aviation Medicine. The first of these involves the collection of critical incidents relating to near accidents in flying reported by Vasilas, Fitzpatrick, DuBois, and Youtz (63). More than 1,700 critical incidents

were collected from pilots and other aircrew members by procedures developed for this study. These incidents pointed to possible improvements in training job design and equipment design as well as in operating procedures.

The second of these studies was specifically concerned with the effect of the age of pilots and other crew members on aircrew operations. This study was reported by Shriver (56), and included tentative suggestions regarding various modifications in operating procedures.

The third study in this series, reported by Goodenough and Suttell (26), involved the collection of critical incidents regarding the impairment of human efficiency in emergency operations. These incidents provide a detailed statement of both the types of stresses that impair performance and the types of performance that are impaired under specific conditions. More than 2,000 critical incidents were collected in which impairment in performance on operational assignments was observed. These incidents were collected in Alaska and the Far East as well as in operational commands in the United States. This report contains suggestions for improving operations in emergency situations.

Equipment design. An application closely related to that just discussed involves the collection of critical incidents to improve the design of equipment. Reports of specific incidents from the field have always been a basis for equipment modifications. The critical incident technique facilitates the collection and processing of this type of information. Too often in the past action was taken on the basis of informal reports from operating personnel. The collection of large numbers of critical incidents

representative of operating experience provides a sound basis for modifying existing equipment and designing new models.

In the study by Fitts and Jones (12), mentioned above, which was carried out at the Aero-Medical Laboratory, 270 critical incidents relating to errors in reading and interpreting aircraft instruments were collected and analyzed. These led to a number of specific suggestions regarding modifications in instrument displays.

Other recent studies conducted at the American Institute for Research have used data from the critical incident technique along with other sources to develop procedures for designing jobs. The reports on these projects are classified for military security reasons.

Other projects at the American Institute for Research have used the critical incident technique as a supplemental procedure for task analysis of equipment in the design stage of development (9, 10, 34, 35, 39). These procedures have been found very effective when used by psychologists working closely with engineers on the preparation of design specifications for new equipment.

Motivation and leadership. The study of attitudes has been somewhat limited and difficult to interpret because of the almost exclusive reliance on verbal statements of opinions and preferences. The critical incident technique has been applied in a few instances to gather factual data regarding specific actions involving decisions and choices. These studies suggest that critical incidents of this type may be a very valuable supplementary tool for the study of attitudes.

A recent study carried out by Preston of the American Institute for

Research for the Air Force's Human Resources Research Center (53) used critical incidents as a basis for studying decisions of airmen to re-enlist in the Air Force. It is believed that these specific incidents provide valuable information not contained in studies utilizing only data on opinions.

A series of reports by Ruch (55) contains critical incidents on combat leadership collected from senior officers in the Far East Air Forces. These incidents provide a factual basis for the study of motivation and leadership of Air Force personnel engaged in combat operations.

Counseling and psychotherapy. Another field in which current techniques emphasize over-all impressions, opinions, and reports of single cases is counseling and psychotherapy. There appears to be a trend, however, in this field toward emphasizing the collection of factual incidents. This suggests that the critical incident technique may be useful in this area also.

Exploratory work has recently been done at the University of Pittsburgh with the critical incident technique to establish areas of change accompanying psychotherapy. A series of three master's theses were carried out by Speth, Goldfarb, and Mellett (25, 38, 60). They collected 243 critical incidents from 11 psychotherapists. These incidents were collected about patients who had shown improvement and were replies to the question, "What did the patient do that was indicative of improvement?" Although these studies were primarily exploratory in nature, the tentative finding that different therapists stress different criteria of improvement and nonimprovement suggests that the critical incident approach may be of use not only in developing objective measures of improvement but also in experimental studies of the types of improvement resulting from the therapists' use of specific procedures.

A somewhat related type of study initiated by Diederich and reported by Allen (1) describes the use of the technique to obtain critical incidents from students reporting things that caused them to like a fellow high school student either more or less than before. This study is being continued to provide the basis for tests of specific value areas. An incidental finding of the study was that when an example of the kind of incident desired was shown on the form, 53 per cent of the positive and 23 per cent of the negative behaviors reported were in the same category as the example given.

SUMMARY AND CONCLUSIONS

This review has described the development of a method of studying activity requirements called the critical incident technique. The technique grew out of studies carried out in the Aviation Psychology Program of the Army Air Forces in World War II. The success of the method in analyzing such activities as combat leadership and disorientation in pilots resulted in its extension and further development after the war. This developmental work has been carried out primarily at the American Institute for Research and the University of Pittsburgh. The reports of this work are reviewed briefly.

The five steps included in the critical incident procedure as most commonly used at the present time are discussed. These are as follows: (a) Determination of the general aim of the activity. This general aim should be a brief statement obtained from the authorities in the field

which expresses in simple terms those objectives to which most people would agree. (b) Development of plans and specifications for collecting factual incidents regarding the activity. The instructions to the persons who are to report their observations need to be as specific as possible with respect to the standards to be used in evaluating and classifying the behavior observed. (c) Collection of the data. The incident may be reported in an interview or written up by the observer himself. In either case it is essential that the reporting be objective and include all relevant details. (d) Analysis of the data. The purpose of this analysis is to summarize and describe the data in an efficient manner so that it can be effectively used for various practical purposes. It is not usually possible to obtain as much objectivity in this step as in the preceding one. (e) Interpretation and reporting of the statement of the requirements of the activity. The possible biases and implications of decisions and procedures made in each of the four previous steps should be clearly reported. The research worker is responsible for pointing out not only the limitations but also the degree of credibility and the value of the final results obtained.

It should be noted that the critical incident technique is very flexible and the principles underlying it have many types of applications. Its two basic principles may be summarized as follows: (a) reporting of facts regarding behavior is preferable to the collection of interpretations, ratings, and opinions based on general impressions; (b) reporting should be limited to those behaviors which, according to competent observers, make a significant contribution to the activity.

It should be emphasized that critical incidents represent only raw data and do not automatically provide solutions to problems. However, a procedure which assists in collecting representative samples of data that are directly relevant to important problems such as establishing standards, determining requirements, or evaluating results should have wide applicability.

The applications of the critical incident technique which have been made to date are discussed under the following nine headings: (a) Measures of typical performance (criteria); (b) measures of proficiency (standard samples); (c) training; (d) selection and classification; (e) job design and purification; (f) operating procedures; (g) equipment design; (h) motivation and leadership (attitudes); (i) counseling and psychotherapy.

In summary, the critical incident technique, rather than collecting opinions, hunches, and estimates, obtains a record of specific behaviors from those in the best position to make the necessary observations and evaluations. The collection and tabulation of these observations make it possible to formulate the critical requirements of an activity. A list of critical behaviors provides a sound basis for making inferences as to requirements in terms of aptitudes, training, and other characteristics. It is believed that progress has been made in the development of procedures for determining activity requirements with objectivity and precision in terms of well-defined and general psychological categories. Much remains to be done. It is hoped that the critical incident technique and related developments will provide a stable foundation for procedures in many areas of psychology.

REFERENCES

1. ALLEN, C. D. Critical requirements in in-
   _____ behavior. Unpublished sen-
   ior thesis, Princeton Univer., 1950.

2. AMERICAN INSTITUTE FOR RESEARCH.
   The applicant inventory. Pittsburgh:
   American Institute for Research, 1954.

3. AMERICAN PSYCHOLOGICAL ASSOCIATION.
   COMMITTEE ON ETHICAL STANDARDS
   FOR PSYCHOLOGY. *Ethical standards of
   psychologists.* Washington, D. C.:
   Amer. Psychol. Ass., 1953.

4. BARNHART, R. E. The critical require-
   ments for school board membership
   based upon an analysis of critical inci-
   dents. Unpublished doctor's disserta-
   tion, Indiana Univer., 1952.

5. COLLINS, MARJORIE G. Selected methods
   applied to the evaluation of a parent
   training course in child management.
   Unpublished doctor's dissertation,
   Univer. of Pittsburgh, 1954.

6. DOMAS, S. J. *Report of an exploratory
   study of teacher competence.* Cambridge:
   New England School Development
   Council, 1950.

7. EILBERT, L. R. A study of emotional im-
   maturity utilizing the critical incident
   technique. *Univer. Pittsburgh Bull.*,
   1953, 49, 199–204.

8. EMONS, V. L'analyse de la fonction de
   vendeuse de grand magasin par la
   méthode des exigencies critiques. Un-
   published doctor's dissertation, Univer.
   of Liége, Belgium, 1952.

9. ERICKSEN, S. C. Development of an ob-
   jective proficiency check for private pi-
   lot certification. Washington: Civil
   Aeronautics Administration, 1951.
   (*Div. of Res., Rep.* No. 95.)

10. ERICKSEN, S. C. Development of a light
    plane proficiency check to predict mili-
    tary flying success. *USAF, Hum. Re-
    sour. Res. Cent. Tech. Rep.*, 1952, No.
    52–6.

11. FINKLE, R. B. A study of the critical re-
    quirements of foremanship. *Univer.
    Pittsburgh Bull.*, 1950, 46, 291–297. (Ab-
    stract)

12. FITTS, P. M., & JONES, R. E. Psychologi-
    cal aspects of instrument display. 1.
    Analysis of 270 "pilot error" experi-
    ences in reading and interpreting air-
    craft instruments. Dayton, O.: U. S.
    Air Force, Air Materiel Command,
    Wright-Patterson Air Force Base, 1947.
    (*Mem. Rep.* TSEAA-694-12A.)

13. FLANAGAN, J. C. The aviation psychology
    program in the Army Air Forces. Wash-
    ington: U. S. Government Printing Of-
    fice, 1947. (*AAF Aviat. Psychol. Pro-
    gram Res. Rep.* No. 1.)

14. FLANAGAN, J. C. A new approach to eval-
    uating personnel. *Personnel*, 1949, 26,
    35–42.

15. FLANAGAN, J. C. Critical requirements: a
    new approach to employee evaluation.
    Personnel Psychol., 1949, 2, 419–425.

16. FLANAGAN, J. C. Use of comprehensive
    rationales. *Educ. psychol. Measmt*,
    1951, 11, 151–155.

17. FLANAGAN, J. C. Principles and proce-
    dures in evaluating performance. *Per-
    sonnel*, 1952, 28, 373–386.

18. FLANAGAN, J. C. *Flanagan aptitude classi-
    fication tests.* Chicago: Science Research
    Associates, 1953.

19. FLANAGAN, J. C. Improving personnel se-
    lection. *Publ. Personnel Rev.*, 1953, 14,
    107–112.

20. FLANAGAN, J. C., et al. *Critical require-
    ments for research personnel.* Pitts-
    burgh: American Institute for Research,
    1949.

21. FLANAGAN, J. C., MILLER, R. B., BURNS,
    R. K., HENDRIX, A. A., STEWART, B.,
    PRESTON, H. O., & WEST, E. D. *The
    performance record for hourly employes.*
    Chicago: Science Research Associates,
    1953.

22. FLANAGAN, J. C., MILLER, R. B., BURNS,
    R. K., HENDRIX, A. A., STEWART, B.,
    PRESTON, H. O., & WEST, E. D. *The
    performance record for non-supervisory
    salaried employes.* Chicago: Science Re-
    search Associates, 1953.

23. FLANAGAN, J. C., MILLER, R. B., BURNS,
    R. K., HENDRIX, A. A., STEWART, B.,
    PRESTON, H. O., & WEST, E. D. *The
    performance record for foremen and su-
    pervisors.* Chicago: Science Research
    Associates, 1953.

24. FOLLEY, J. D., JR. Development of a list
    of critical requirements for retail sales
    personnel from the standpoint of cus-
    tomer satisfaction. Unpublished mas-
    ter's thesis, Univer. of Pittsburgh, 1953.

25. GOLDFARB, A. Use of the critical incident
    technique to establish areas of change
    accompanying psychotherapy: II. Re-
    lationship to diagnostic group. Unpub-
    lished master's thesis, Univer. of Pitts-
    burgh, 1952.

26. GOODENOUGH, D. R., & SUTTELL, BAR-
    BARA J. *The nature and extent of impair-
    ment of human efficiency in emergency op-
    erations: the field study.* Pittsburgh:

American Institute for Research, 1952.

27. GORDON, T. The airline pilot: a survey of the critical requirements of his job and of pilot evaluation and selection procedures. Washington: Civil Aeronautics Administration, 1947. (*Div. of Res., Rep.* No. 73.)

28. GORDON, T. The development of a standard flight-check for the airline transport rating based on the critical requirements of the airline pilot's job. Washington: Civil Aeronautics Administration, 1949. (*Div. of Res., Rep.* No. 85.)

29. GORDON, T. The development of a method of evaluating flying skill. *Personnel Psychol.*, 1950, 3, 71–84.

30. HAHN, C. P. Measurement of individual differences with respect to critical job requirements. *USAF Sch. Aviat. Med.*, 1954, Proj. No. 21-29-014, No. 2.

31. HOBBS, N. The development of a code of ethical standards for psychology. *Amer. Psychologist*, 1948, 3, 80–84.

32. JENSEN, A. C. Determining critical requirements for teachers. *J. exp. Educ.*, 1951, 20, 79–86.

33. KONIGSBURG, D. The development and preliminary evaluation of an instructor check list based on the critical incident technique. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1954.

34. KRUMM, R. L. Critical requirements of pilot instructors. *USAF, Hum. Resour. Res. Cent., Tech. Rep.*, 1952, No. 52-1.

35. KRUMM, R. L. *Development of a measure of pilot instructor proficiency based on the critical requirements of the instructor's job.* Pittsburgh: American Institute for Research, 1953.

36. MARLEY, F. W. Individual differences in critical aircrew elements. I. The determination of critical proficiency requirements for B-29 combat crews. *USAF Sch. Aviat. Med.*, 1952, Proj. No. 21-29-014, No. 1.

37. MARLEY, F. W. The development of performance flight checks for B-29 combat crews. *USAF, Hum. Resour. Res. Lab.*, 1952, Rep. No. 19.

38. MELLETT, T. P. Use of the critical incident technique to establish areas of change accompanying psychotherapy: III. Differences among therapists. Unpublished master's thesis, Univer. of Pittsburgh, 1952.

39. MILLER, G. S. *Development of B-36 bombardment objective flight checks and proficiency measures for use with reconnaissance crews.* Pittsburgh: American Institute for Research, 1953.

40. MILLER, N. E. Psychological research on pilot training. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program. Res. Rep.* No. 8.)

41. MILLER, R. B. A ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯ ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯ *Res. Cent. Dev.* ⋯ 1953, N. 51-1.

42. MILLER, R. B. A ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯ human ⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯⋯ for training equipment. Dayton, O.: U. S. Air Force, Air Development Center, Wright-Patterson Air Force Base, 1953. (*Tech. Rep.* 53-135.)

43. MILLER, R. B. Handbook on training and training equipment design. Dayton, O.: U. S. Air Force, Air Development Center, Wright-Patterson Air Force Base, 1953. (*Tech. Rep.* 53-136.)

44. MILLER, R. B. A method for man-machine task analysis. Dayton, O.: U. S. Air Force, Air Development Center, Wright-Patterson Air Force Base, 1953. (*Tech. Rep.* 53-137.)

45. MILLER, R. B. Human engineering design schedule for training equipment. Dayton, O.: U. S. Air Force, Air Development Center, Wright-Patterson Air Force Base, 1953. (*Tech. Rep.* 53-138.)

46. MILLER, R. B., & FLANAGAN, J. C. The performance record: an objective merit-rating procedure for industry. *Amer. Psychologist*, 1950, 5, 331–332. (Abstract)

47. MILLER, R. B., & FOLLEY, J. D., JR. *The validity of maintenance job analysis from the prototype of an electronic equipment.* Pittsburgh: American Institute for Research, 1952.

48. NAGAY, J. A. The development of a procedure for evaluating the proficiency of air route traffic controllers. Washington: Civil Aeronautics Administration, 1949. (*Div. of Res., Rep.* No. 83.)

49. NAGAY, J. A. The airline tryout of the standard flight-check for the airline transport rating. Washington: Civil Aeronautics Administration, 1949. (*Div. of Res., Rep.* No. 88.)

50. NEVINS, CHARLOTTE I. An analysis of reasons for the success or failure of bookkeepers in sales companies. Unpublished master's thesis, Univer. of Pittsburgh, 1949.

51. O'DONNELL, R. J. The development and evaluation of a test for predicting dental student performance. *Univer. Pittsburgh Bull.*, 1953, 49, 240–245. (Abstract)

52. PRESTON, H. O. *The development of a pro-*

cedure for evaluating officers in the United States Air Force. Pittsburgh: American Institute for Research, 1948.

53. PRESTON, H. O. *Events affecting reenlistment decisions.* Pittsburgh: American Institute for Research, 1953.

54. RONAN, W. W. *Training for emergency procedures in multi-engine aircraft.* Pittsburgh: American Institute for Research, 1953.

55. RUCH, F. L. Incidents of leadership in combat. *USAF, Hum. Resour. Res. Inst., Res. Memo,* 1953, No. 3.

56. SHRIVER, BEATRICE M. Age and behavior: a study of the effects of aging on aircrew performance. *USAF Sch. Aviat. Med.,* 1953, Proj. No. 21-0202-0005, No. 3.

57. SIVY, J., & LANGE, C. J. Development of an objective form of the Leaders Reaction Test. *AGO, Personnel Res. Branch, PRS Rep.* 1952, No. 930.

58. SMIT, JO ANNE. A study of the critical requirements for instructors of general psychology courses. *Univer. Pittsburgh Bull.,* 1952, 48, 279–284. (Abstract)

59. SMITH, R. G., JR., & STAUDOHAR, F. T. Technical requirements of basic training tactical instructors. *USAF, Hum. Resour. Res. Cent., Tech. Rep.,* 1954, No. 54-7.

60. SPETH, E. W. The use of the critical incident technique to establish areas of change accompanying psychotherapy: I. Function of age and education. Unpublished master's thesis, Univer. of Pittsburgh, 1952.

61. SUTTELL, BARBARA J. *Development of a situation test to measure non-intellectual officer qualities.* Pittsburgh: American Institute for Research, 1953.

62. TAYLOR, M. V., JR. *The development of aircrew job element aptitude tests.* Pittsburgh: American Institute for Research, 1953.

63. VASILAS, J. N., FITZPATRICK, R., DUBOIS, P. H., & YOUTZ, R. P. Human factors in near accidents. *USAF Sch. Aviat. Med.,* 1953, Proj. No. 21-1207-0001, No. 1.

64. WAGER, C. F., & SHARON, M. I. Defining job requirements in terms of behavior. *Personnel Admin.,* 1951, 14, 18–25.

65. WAGNER, R. F. A group situation compared with individual interviews for securing personnel information. *Personnel Psychol.,* 1948, 1, 93–107.

66. WAGNER, R. F. A study of the critical requirements for dentists. *Univer. Pittsburgh Bull.,* 1950, 46, 331–339. (Abstract)

67. WAGNER, R. F. Development of standardized procedures for defining the requirements of aircrew jobs in terms of testable traits. *USAF Sch. Aviat. Med.,* 1951, Proj. No. 21-29-010, No. 1.

68. WAGNER, R. F. Using critical incidents to determine selection test weights. *Personnel Psychol.,* 1951, 4, 373–381.

69. WEISLOGEL, MARY H. *Development of a test for selecting research personnel.* Pittsburgh: American Institute for Research, 1950.

70. WEISLOGEL, MARY H. *Procedures for evaluating research personnel with a performance record of critical incidents.* Pittsburgh: American Institute for Research, 1950.

71. WEISLOGEL, MARY H. *The development of tests for evaluating research proficiency in physics and chemistry.* Pittsburgh: American Institute for Research, 1951.

72. WEISLOGEL, R. L. Critical requirements for life insurance agency heads. *Univer. Pittsburgh Bull.,* 1952, 48, 300–305. (Abstract)

73. WEISLOGEL, R. L., & SCHWARZ, P. A. Some practical and theoretical problems in situational testing. Unpublished manuscript, Pittsburgh, American Institute for Research, 1953.

74. WICKERT, F. Psychological research on problems of redistribution. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 14.)

THE VALIDITY OF SOME PSYCHOLOGICAL TESTS OF BRAIN DAMAGE[1]

AUBREY J. YATES

Institute of Psychiatry, Maudsley Hospital, London

It is customary to distinguish two main types of validity; these may be named *higher-order* (internal) and *lower-order* (external) validity. Internal validation consists essentially in measuring the correlation between different tests supposed to measure the same variable. Consideration of this method of validation will be reserved until later, since the tests to be evaluated below have not generally been validated in this way, and the use of this method is comparatively rare. Lower-order validity implies the use of an external criterion against which the test is validated. In the development of a test of brain damage, for instance, the external criterion would consist of a number of clinical groups, such as a brain-damaged group of patients, a group of psychiatric (functional) patients without brain damage, and a group of normal controls; a valid test of brain damage would then be expected to distinguish patients in the brain-damaged group from those in the other two groups. The difficulty in using this method of validation lies, of course, in the fact that the criterion itself is in need of validation. It has been shown by Ash (5) and others that the reliability of psychi-

atric diagnosis is so low that it is difficult to rely on such classifications in the development of tests. Even when very broad groups are used, the degree of error may be considerable. It seems necessary to use the classification system given above, however, because this is the procedure adopted by the authors of most of the tests to be discussed, because the groups may be more reliably discriminated than in most cases, and because, in the case of the brain-damaged group at least, independent confirmation may be forthcoming in the shape of neurological signs, post-mortem examination, etc.

It is assumed, then, that it is possible to isolate such broad groups for the purposes of research; however, there remain other conditions which must be fulfilled before the validity of a given test can be discussed. In general, these may be summarized as follows:

a. The test should present data for adequate samples of the above-mentioned three groups (brain-damaged, functional, and normal).

b. The data should be presented in such a form as to enable the clinician using the test to estimate the degree of possible error when assigning a patient to any one of the groups. Such data may be stated in three different ways. First, the optimum cutoff point may be given; this is the point at which it is possible to identify as many brain-damaged patients as possible and at the same time misclassify as few functionals and normals as possible. (If the distributions are normal, this can be stated simply in terms of the mean and standard deviation.) Second, the point beyond which no normals or functionals of the samples fall should be given. Third, the point beyond which no brain-damaged patient of the sample falls should also be given.

c. The reliability of the results should be verified by applying the test to new groups that are independent of the original criterion groups. Alternatively (or, preferably, in addi-

tion), the findings should be confirmed by another worker in another hospital.

d. The influence of various factors, such as age, sex, and intelligence, and any special factors (such as visual acuity when perceptual tests are used, and motor coordination when mechanical tests are used) should be controlled.

It will be the purpose of this paper to show that most of the tests purporting to be measures of brain damage do not meet the conditions set out above and that, therefore, their validity either cannot be considered as established or cannot even be evaluated. It may be pointed out, however, that, even if these conditions were adequately met, no general judgment can be made about the validity of a particular test. Whether a given test is considered to be a valid measure of brain damage, or whether it is not, will depend on many factors. The most important of these will be whether or not the test works in practice. Because a test is usually standardized on relatively pure groups, the discriminating power must necessarily drop when the test is used clinically. Again, precise confidence limits are difficult to apply. Thus, a test that identifies only 20 per cent of brain-damaged patients admitted to a hospital may be a very useful clinical instrument if these patients are not identifiable at the time of testing by any other means. On the other hand, a test that identifies 60 per cent of all brain-damaged patients may not be very useful clinically because most of the patients it identifies are obvious cases of brain damage and can be detected by simpler means. In some instances, it may be possible to use a cutoff point in such a way that a patient falling above it is unequivocally identified as brain-damaged, while, if he falls below it, he (being brain-damaged) is not misclassified; but the question remains an open one, no state-ment being made about the patient. However, when considering the standardization of a test, one may reasonably demand a low percentage of misclassification, because the groups are carefully chosen for their clinical differences. The various tests of brain damage will now be considered to see how far they fulfill the criteria already laid down.

The tests will be considered under two broad headings—those tests which employ qualitative methods and those which employ quantitative methods. Those tests using quantitative methods will, in turn, be grouped into those utilizing the concept of deterioration and those measuring perceptual or motor functions. Such a division is, of course, quite arbitrary, and is dictated by convenience.

QUALITATIVE TESTS OF BRAIN DAMAGE

The criteria laid down above presuppose adequate statistical treatment of data. However, such treatment is almost entirely lacking in one of the most widely used and reputable batteries of tests of brain damage—the tests of abstract concept formation developed by Goldstein and Scheerer (**22**) and their colleagues. These tests are so well known that it is unnecessary to describe them. The basic criticism to be made of these tests is not that they are invalid, but that there is no basis for a discussion of their validity. The tests are unique in diagnostic psychological testing in providing no quantitative data on the subjects used and providing no percentage of incorrect diagnoses, in ignoring the effects of age and intelligence level completely, and in assuming that the performance of normal people of average intelligence will be without error. Under these circum-

stances, it is clear that the validity of the tests is something private to each individual user. It is possible that, in the hands of a skilled clinician, the tests have a high validity; but, even in the case of Goldstein himself, no information is available about the number of times he has been in error. Two further points may be made: First, there has never been any clear agreement concerning what the tests measure. Goldstein believes that they measure the ability to abstract, and that brain injury leads to a defect in assuming the abstract attitude, which is revealed in test performance; on the other hand, Hutton (36) believes that failure on the Block Design test is due to overabstraction, not failure of abstraction. Second, it is significant that the Goldstein-Scheerer tests have been used to determine the presence or absence of schizophrenic thought disorder, and that other tests, developed specifically to test for schizophrenia, bear a strong resemblance to the Goldstein tests, e.g., the Hanfmann-Kasanin Test of Concept Formation. Little work has been done to differentiate between the performance of schizophrenics and brain-damaged patients on these tests; hence, they are very difficult to apply clinically.

That the necessary quantification would be possible may be easily inferred from examination of Goldstein's manual, where many possible dimensions of measurement are indicated in qualitative form. A beginning has, in fact, been made. Boyd (7), using 54 normal hospitalized subjects, set up quantitative norms for the Block Design test by giving weights to the various steps utilized by Goldstein. When he divided his group according to Wechsler IQ, he found that perfect performance would indeed be expected, provided the IQ

was 100 or more. As the IQ dropped, however, so did the score on the Block Design test, from a mean score of 120 ($SD = 0$) for IQ 100–109 to a mean score of only 99.5 ($SD = 34.5$) for IQ 66–79. He was able to give limits within which a person of a given IQ level would be expected to fall. He thus demonstrated that intelligence is an important factor in performance on these tests; he also found that psychotics tended to do worse than brain-damaged patients (though his numbers were very small).

Lidz, Gay, and Tietze (41), using the scoring system developed by Kohs, found that the Kohs test discriminated significantly between 21 organics with deterioration, and 15 nondeteriorated schizophrenics (with a misclassification of only 2/36). However, there was no control for age, the difference between the two groups being certainly significant. Another attempt to quantify the Block Design test was reported by Armitage (4). Using three groups of subjects—normal, neurotic, and brain-damaged—he calculated the percentage of controls and organics requiring assistance at each of four steps. He concluded that the quantitative results were disappointing from the standpoint of a screening device; so, to determine whether the discriminatory ability of the test could be increased, he made a multiple approach. This involved the computation of such variables as the time taken to complete each design, the number of incorrect moves, the order or sequence of placing the blocks, and the number of correct first block placements. None of these methods proved successful.

Tooth (67) gave a series of tests, including the Kohs Block test and the Weigl Color Form Sorting Test, to

100 cooperative naval officers and ratings admitted to a naval hospital with a history of an injury to the head and a diagnosis of postconcussional state. As controls, he used 50 convalescent patients in the surgical wards of the same hospital and 50 neurotics carefully chosen from 2,000 cases seen in a naval psychiatric clinic. After careful statistical analysis of the results, he concluded that the method did not give sufficient quantitative discrimination between head injury cases and normals to be of much practical importance in the assessment of the individual case. Furthermore, in these two tests, a difference of as great a magnitude was found between the normal controls on the one hand and a series of neurotic patients in whom no organic condition was known to exist on the other.

With respect to object sorting, Halstead (27), using groups of cases with lesions in various areas of the brain, compared them with a normal group on an object-sorting test, larger in content and presented under conditions somewhat different from those obtaining in the Goldstein test. The sorting behavior was analyzed according to a number of criteria, including the percentage of objects grouped, the total percentage of objects recalled after five minutes, and the percentage of objects grouped that were recalled after the same interval. With respect to these three variables, for 11 normals and 11 cases of frontal lobe injury there was no overlap in the scores of the two groups on the first and third variables, and only two brain-damaged cases exceeded the lowest normal score on the second variable. On the other hand, cases with lesions in other parts of the brain showed considerable overlap with normals. Halstead's results confirmed to some

extent Goldstein's hypothesis that performance as a whole on this test tends to be lowered in cases of frontal lobe lesion. He also offered some evidence for the hypothesis that patients with injury to the frontal lobes cannot sort according to an abstract principle. In terms of the number of abstract groupings produced, there was again no overlap between the normals and the cases of frontal lobe injury.

A rather radical and important departure from the usual procedure in the Block Design test has been made by Grassi (23, 24) in his Block Substitution Test for measuring organic brain pathology. Essentially, the main innovations are that the patient copies not a drawing but a set of blocks, and an attempt is made to measure both concrete and abstract reproduction at two levels. The test consists of five designs or models, each of which is reproduced by the patient at four levels of increasing complexity. In the simple concrete task, the patient copies only the top side of the model. In the simple abstract task, he copies the top side again, but this time using different colors from those of the model. In the complex concrete task, the patient has to copy the model correctly with respect to all six sides, while in the complex abstract task, he copies the complete model again, but this time using different colors from those of the model. The scoring system is simple, accuracy and time being rewarded; the maximum score possible is 30. The standardization groups are unusually large for this kind of work, consisting of 86 normals, 86 schizophrenics (equally divided into deteriorated and nondeteriorated cases), 72 organics, and 30 postlobotomy cases. Grassi states that no influence of age, sex, or other factors

was found, while test-retest relia-
bility was high (+.85). Means and
ranges are given for the four groups
and show that (using a cutoff point
of 16) there was no overlap between
the two organic and the two nonor-
ganic groups. If this finding can be
confirmed, it is clearly of great impor-
tance. There is some doubt, however,
about the accuracy of Grassi's fig-
ures.   For example, although the
ranges of the organic and nonorganic
groups do not overlap, Grassi reports
in another section that six schizo-
phrenics scored below 16 and were
not included in the standardization
data; also, that eight organics scored
"slightly above" 16 and were cor-
rectly identified by qualitative exami
nation.   However, the misclassifica-
tion is sufficiently small to make fur-
ther research on this test imperative.

Grassi's (24) attitude toward his
own statistics is difficult to under-
stand. He writes: "It cannot be too
strongly stressed that test scores
taken alone will lead to incorrect
conclusions. The score is intended as
a *guide*, not as the *sole* criterion for
the final conclusion. Behavior cannot
be too greatly emphasized and should
be, in most cases, the basis for diag-
nosis.   Test score and intellectual
level are supporting factors.   They
must not, *ever*, be the sole basis for
final classification" (24, p. 63, italics
his).   Two comments are rele-
vant here.   In the first place, for
the standardization groups at least,
since there is no overlap, it is hard to
see how a quantitative score could
lead to a false diagnosis of brain dam-
age. Second, a subjective statement
that the patient's behavior was quali-
tatively such and such is in fact a
preliminary quantitative statement
of behavior, albeit in a crude and
inaccurate way. The aim of any test
constructor must surely be to quan-

tify as soon as possible his initially
crude and unrefined observations
and thus set up proper control. Why
Grassi, having done this, should then
turn his back on it, is difficult to see.
The consequences, however, are seri-
ous. Although Grassi himself states
that intelligence can influence per-
formance, no adequate information is
given. This would have been espe-
cially valuable in view of the rela-
tively large numbers of subjects used.
In spite of these criticisms, the Grassi
test represents an important advance
over the original Goldstein tests.

Clinical use of the Color Form test
of the Goldstein battery has sug-
gested that it is too easy to be useful
in the detection of brain damage
except in severe cases. This was con-
firmed by McFie and Piercy (43) who
used 55 cases of known brain damage
and scored the test simply as "pass"
or "fail" on the basis of the patient's
ability to sort the pieces both by form
and by color. They found that only
9 of the 55 patients failed the test.
An important modification of the
Color Form test has, however, been
published by Scheerer (60). Three
large model figures—a circle, a tri-
angle, and a square—are used, and
there are 12 test figures, four being
subsumed under the concept of
roundness, four under the concept of
triangularity, and four under that of
squareness. The three large figures
lie on the table as model figures. The
subject is shown each of the remain-
ing figures in turn and asked to indi-
cate with which model figure it be-
longs. Each figure is removed from
the subject's sight before the next
one is presented. A simple, objective
system of scoring has been devised,
failure being counted when the sub-
ject incorrectly assigns a figure. If
this happens, a series of three "helps"
is given, similar in nature to the suc-

cessive steps on the Block Design test. The test was standardized on four groups of subjects. The first group consisted of 44 college students; the second group was made up of 20 noncollege students; the third group consisted of 20 brain-damaged patients; and the fourth group included 20 retarded high school pupils who were used as controls for those brain-damaged patients of low intelligence. Scheerer found that the percentage of subjects needing the three "helps" fell rapidly after the first help for the three non-brain-damaged groups, but that all the brain-damaged patients needed the third help and 75 per cent of them failed all three helps; no subjects in the other three groups failed all three helps.

This study is of importance, not only for its success in identifying brain-damaged patients, but in representing one of the few serious attempts at quantification of the Goldstein-Scheerer tests. However, some criticisms are possible. The instructions contained in the original article are inadequate for clinical use—for instance, it is not made clear what happens if the subject makes more than one error after the first help has been given; the test objects are presented in a fixed, not random, order, so that a pattern of response may be set up which masks inability to do the test. In this connection it may be noted that the figures given by Scheerer indicate that, compared with the noncollege group, fewer of the retarded group needed the third help, and that, compared with the college group, fewer retarded needed the second help. Again, the test has not been given to a group of functional patients, so the degree of overlap between such a group and the brain-damaged group is not known. As far as is known, no subsequent in-

formation has been made available on this test, and no critical studies have been published.

Zangwill (in Buros, 10) has criticized the Goldstein tests on the general grounds that they are qualitative; that some psychotics show impairment; that some brain-damaged patients with traumatic injuries behave essentially normally on the tests; and that gross impairment on the tests due to a specific lesion, may be present in the absence of any occupational or social difficulties. It may be added that no information is available concerning the factorial composition of the tests.

QUANTITATIVE TESTS OF BRAIN DAMAGE

Tests Employing the Concept of Deterioration

The tests considered under this heading are based on the principle that brain damage leads to deterioration of an irreversible nature. This deterioration usually is contrasted with that found in the functional disorders, wherein deterioration is considered to be more apparent than real, and to be due to inattention, the assumption being that the patient retains normal ability if he could be made to perform at his best level. It is not proposed to discuss the concept of deterioration as such, and these tests will be criticized only in so far as they claim to distinguish brain-damaged patients from other groups.

The Hunt-Minnesota Test for Organic Brain Damage (34) consists of three major divisions—a vocabulary test, which is relatively insensitive to brain damage; a group of tests sensitive to deterioration; and a group of interpolated tests. The subject's Stanford-Binet vocabulary score, in relation to his age, determines the

score level at which he is expected to perform the more sensitive tests. The deterioration tests, consisting of pairs of words and of designs that the subject is expected to associate and later recall and recognize, determine the level at which he is actually functioning. The amount of discrepancy between the subject's expected score and the score he actually makes on the word and design associations (when corrected for age) is the basis for the diagnosis of brain damage. The discrepancies are indicated by T scores; those T scores which fall higher than a certain critical point are considered to be indicative of brain damage. Originally, this critical T score was fixed at 68; later it was reduced to 66. The test may be given in either a short or long form.

Two groups of subjects were used in the development of the test: 33 patients suffering from brain damage (all but three were cases of diffuse damage); and 41 controls, consisting of 15 neurotics, 11 normals, 6 psychotics, and 8 nonpsychiatric patients. All subjects had to be able to read and speak English, had attended school to third grade, were adequately cooperative and attentive, had adequate muscular coordination and sensory acuity, and were between the ages of 16 and 70. For discriminative purposes, 25 persons equated for vocabulary and age were drawn from each group. Using a cutting point of 68, Hunt found that only one of the 50 standardization subjects was misclassified, and that, of the remaining 24 subjects, only three were misclassified. (He ignored three subjects who scored 68.)

Hunt's use of the interpolated tests (tests of concentration and attention) was based on the following assumptions: deteriorated brain-damaged patients would fail the learning tests

but not the interpolated tests; functional, "apparently deteriorated" patients would succeed on both learning and interpolated tests; functional, "genuinely deteriorated" patients would fail both learning and interpolated tests. Implicitly, therefore, failure on the interpolated tests was to be utilized as a measure of whether a functional patient was really, or was only apparently, deteriorated. In fact, this ingenious hypothesis was not directly tested by Hunt; his only finding was that both of the groups he used were not differentiated by the interpolated tests.

While, in its construction and standardization, this test fulfills to some extent the first two criteria laid down earlier in this paper, the results reported by Hunt have not been confirmed by other workers, while important reservations in the use of the Hunt-Minnesota vocabulary test are necessary.

The discrimination between normals and organics has not been found by other investigators. Thus, Aita and his associates (1), using groups already referred to in the study by Armitage (4), found that 10 per cent of their control subjects obtained a T score greater than 90, but that only 2.2 per cent of their brain-damaged subjects obtained a score as high as this. On the other hand, 41.3 per cent of their brain-damaged subjects obtained a score that was in the normal range, i.e., below 70. Many of the brain-damaged patients actually scored well below the cutoff point used by Hunt, 23.9 per cent of them falling within the range 50–54. When the brain-damaged group was divided into three categories— severe, moderate, and mild—only the mean T score of the severe group was significantly different from that of the controls.

A study by Canter (11) reinforced this finding that the amount of overlap between a brain-damaged group and a normal group may be considerable. For 47 arteriosclerotic patients the mean T score was 69.04 (*SD* = 12.49)—that is, the group mean fell within normal limits. Malamud (45), in a preliminary study, found that six out of ten members of the psychology department at her hospital were suffering from brain damage according to results on this test! She therefore administered the test to 64 subjects who satisfied Hunt's basic criteria. She found that 54.7 per cent of these normal subjects obtained scores indicating organic brain damage. The exact critical score used did not influence the results appreciably. When all vocabulary scores above 33 words or more were arbitrarily reduced to 32, no less than 42.2 per cent of the total group still fell within the pathological range.

In addition to this overlap between organics and normals, it has been shown by Meehl and Jeffery (47) that some functionals at least may show pathological scores on this test. They gave the test to a group of 15 patients with functional depressions, of whom 9 were psychotic and 6 neurotic. The possibility of brain damage was ruled out in all cases, and 11 of the subjects did not fail any of the interpolated tests. The mean T score of the entire group was 70.2 (*SD*=15.41). Using a critical score of 70, they found that one in four functionally depressed patients would obtain an organic score; using a cutoff point of 66, the ratio would be one in three.

We have already seen that Malamud found that a high vocabulary score did not markedly affect the proportion of normals showing a pathological score on this test. However, a study by Juckem and Wold (37) showed that when the test was given to a group of 50 college students who reached Superior Adult III level on the Binet vocabulary test, they obtained a mean T score of 69.6, whereas Hunt's control-group mean was 50. Only 3 per cent of Hunt's control group obtained a score greater than this mean score. Of this college population, no less than 60 per cent had scores above Hunt's critical point. Juckem and Wold conclude that the test yields far too many false positives among persons of high vocabulary level. If a normal T score were to be obtained, the vocabulary level would have to be reduced to 21 words.

These studies show that, in practice, the test has not lived up to the claims made for it by its author, and its usefulness as a test of brain damage is very much in doubt.

The Shipley-Hartford Retreat Scale (64) is a short test of two parts, one consisting of a vocabulary test, which is supposed to hold up with age or illness; the other, of a set of abstract reasoning problems, supposedly sensitive to deterioration. A Conceptual Quotient (CQ) is derived from scores on the two parts of the test and may be defined roughly as Mental Age (Abstractions Test) /Mental Age (Vocabulary Test) ×100. Mental age norms were established on 1,046 young subjects who had been given a group intelligence test. It was assumed that normal persons, regardless of native intelligence, would approach CQ's of 100. This assumption proved to be roughly true within normal limits. A CQ above 90 was considered normal; one between 76 and 89, suspicious; and a CQ below 75, pathological. The test was validated on 171 state hospital cases and 203 private hospital cases.

Shipley and Burlingame (65) claimed that the test would discriminate between normals, schizophrenics, and organics, and it is for this reason that the test is included here. They found that organics did worst, that neurotics and normals fell close together, and that the psychotics came in between.

This test may be criticized severely on many grounds. The standardization data were completely inadequate, only young intelligent normals being used in the construction of the scale. There was no control for age, sex, or intelligence, yet the factor of age is of particular importance in this test. Subsequent research has shown beyond doubt that the test does not discriminate between organic and other groups without considerable overlap.

Thus, Aita and his colleagues (1) found that, of a control group of 61 normal subjects, 47 per cent obtained a CQ that was suspicious or pathological, and 26 per cent of these 61 subjects obtained a definitely pathological score. Of a brain-damaged group of 70 patients, 33 per cent fell within the normal range of scores. Contrary to Shipley's assertion, they found that mild or moderate neurosis lowers the CQ, although they present no figures. Canter (11) used the Shipley-Hartford scale on 47 cases of multiple sclerosis and obtained a mean CQ score of 85.8 ($SD = 14.31$). Despite the unfavorable difference in age between his group (mean age, 32) and the standardization group of Shipley, he found that 50 per cent of his cases obtained CQ's greater than 76, and 40 per cent obtained CQ's greater than 90, indicating considerable overlap with normals.

Magaret and Simpson (44) gave the test to 50 patients consisting of psychotics and neurotics, aged 40–49,

and found a mean CQ of 74 ($SD = 12.7$), with a range of scores from 55 to 105. Of the 50 cases, 29 obtained a CQ lower than 70. Here again, although age would be expected to work in favor of lower scores, the overlap with normals would still be considerable, while the overlap with organics would be even greater. Furthermore, correlations of the CQ with psychiatric ratings of deterioration and with the Wechsler deterioration index were not significantly greater than zero. These results are the more remarkable in that Garfield and Fey (21) showed that, in fact, the CQ did decline steeply, solely as a function of age. Their results otherwise agree with those of Magaret and Simpson: using a group of 100 patients (58 psychotic, 37 neurotic, 13 unclassified), they found a mean CQ of 79.1 for the psychotics and 85.2 for the neurotics.

A further study by Manson and Grayson (46) was even more disturbing. Of Shipley's standardization group, 26 per cent obtained a CQ below 90. Of a group of 1,262 military prisoners examined by these authors, no less than 59.1 per cent obtained a CQ less than 90; 34.6 per cent a CQ less than 80; and 14.2 per cent a CQ less than 70; the mean CQ for the group was only 88.8. They suggested that the reason for this discrepancy was a spuriously high vocabulary score, and showed that substitution of the Army General Classification Scores for the vocabulary scores produced figures very similar to those of Shipley. This confirms the previous findings that spuriously organic scores may be obtained on tests which use vocabulary level as a measure of previous intellectual functioning. Manson and Grayson also report that 45 per cent of severe neurotics scored below 80 on this

test. This was confirmed by the work of Kobler (39), who reported that, of 500 neurotics at an army rehabilitation center for combat fatigue, 78 per cent made CQ's below 90; 26 per cent had CQ's below 70 and thus fell within the organic range.

Fleming (20) reported that 20 depressives obtained a mean CQ of only 73.2. However, the mean age of his group was 57.6 years. Ross and McNaughton (56) used 90 subjects with head injury and found that the results bore no relation to severity of the injury or to EEG or pneumoencephalographic evidence of cerebral damage except in cases of severe injury. No statistical data were given, however.

How far the original faulty standardization of the test and lack of care in its construction have been responsible for the failure of subsequent workers to establish the validity of the test is difficult to estimate. The findings quoted, however, certainly indicate that the test has not been adequately validated.

There have been a number of attempts to develop indices of deterioration by using the Wechsler subtests in various combinations. The rationale of these tests is well known. A ratio is calculated between those tests which are said to hold up with age, and those which do not hold up with age. The underlying assumption is the rather curious one that organic deterioration is similar to the deterioration accompanying age, differing only in its early onset. Wechsler (68) found that the index discriminated between young normals and young brain-damaged patients, the percentage overlap, however, being high. Thus, there was a restriction on the usefulness of the test right from the start. Levi, Oppenheim, and Wechsler (40) specifically claimed that the *DI* was useful, not only in confirming, but also in discovering, organic conditions. They claimed that it would assist in differentiating organic memory impairment from hysterical amnesia; in finding corroborative evidence of organic involvement where clinical and neurological data are not clear-cut; in distinguishing between mental deficiency and mental deterioration; and in differentiating between psychosis with and without organic deterioration. With the exception of the first, these claims have been tested and invariably have been found wanting; and it seems safe to say that the first has not been found wanting only because it has not been tested.

Two important modifications of the Wechsler *DI* are Reynell's index (52), which makes use only of the verbal subtests, and Hewson's deviation ratios (31), which are based on various combinations of the subtests. Reynell's index, however, has been incorrectly criticized by some workers—its intention was not to discriminate brain-damaged cases from others, but to discriminate those brain-damaged cases *with* deterioration from those without (and hence assumed initial knowledge of brain damage).

Many studies have now been published on the *DI*. This paper is concerned only with those investigating the validity of the tests as a measure of brain damage. At least six studies concur in reporting unfavorably on its use. Thus Gutman (25), using 30 organics and 30 controls, found that the Wechsler *DI* correctly identified only 43 per cent of organics, the Reynell index 50 per cent, and the Hewson ratios 60 per cent; whereas the *DI* misclassified 33 per cent of the normal group, the Reynell index 30 per cent, and the Hewson ratios 17 per cent. The three measures agreed in

... the diagnosis of brain damage in only 33 per cent of the cases. Five cases of clinically verified brain damage did not fall in the organic range on any of the tests. Allen (2), using as his criterion of deficit a loss greater than 20 per cent, found that the Wechsler *DI* definitely screened out only 54 per cent of the total study group of 50 patients. Rogers (53) evaluated the *DI* for seven groups ... and found that, using a cutting score of 10 per cent, 75 per cent of subjects will be correctly identified, provided that only the brain-damaged and normal groups are used, but that, when other clinical groups are included, the results are no better than chance. Andersen (3), using 55 male soldiers with definite clinical evidence of brain damage, showed that, when a cutting score of 10 per cent was used, nearly one-third of the total sample fell outside the organic range; yet when a cutting score of 20 per cent was used, nearly two-thirds of the patients fell inside the normal range. He divided his group of patients into those suffering predominantly from injury to the dominant hemisphere and those suffering from injury to the nondominant hemisphere. This did not materially improve the results. Kass (38) gave the test to 18 cases with known organic damage and 12 cases of dubious organic diagnosis, and concluded that the *DI* failed both in detecting and confirming the presence of organic conditions resulting largely from traumatic brain injury. As a percentage-loss method for expressing psychological deficit, it was found inapplicable in two-thirds of his cases. Diers and Brown (15), using 25 cases of multiple sclerosis, concluded that the *DI* was not sensitive enough to be used clinically. Garfield and Fey (21) found that an equal number of psychotic and nonpsy-

chotic patients obtained pathologically high *DI*'s, suggesting that the overlap between organics and functionals would be quite high. Magaret and Simpson (44) found that the *DI* rating did not correlate with the psychiatrist's ratings of degree of deterioration. The only study so far to produce reasonably favorable results with the *DI* was that by McFie and Piercy (43). Using 56 brain-damaged patients and a cutting score of 10 per cent, they were able to identify 43 (71 per cent) of them; using a cutting score of 20 per cent, they identified 37 (66 per cent). No functional patients were tested. In view of the unfavorable results summarized above, it seems clear that indexes of deterioration are of little clinical use in their present form.

Halstead (28), using factorial analysis and various systems of weighting, developed a battery of tests which discriminated at a high level of confidence between normals and patients with lesions of the frontal lobes. The ten tests having the highest *t* value were selected as the basis of an impairment index. In this arrangement, an individual whose scores fell below the criterion scores on all ten of the key tests had an impairment index of 0.0; while, on a simple proportion basis, an individual who satisfied the criterion score on three of ten key tests had an impairment index of 0.3; or on all of the key tests, an index of 1.0. Using a cutting score of three, he was able to identify all 27 cases of frontal lobe injury and 29 out of 30 normals. The impairment index did not discriminate between normals and other cases of brain damage.

Of the ten tests, the Halstead Category Test (28, 29), involving the ability of the subject to "abstract" various organizing principles such as "size," "shape," "color," etc. from a

series of 336 stimulus figures presented visually and serially by means of a multiple-choice projection apparatus, proved particularly successful. Using a cutting score of 70, he correctly identified 27 out of 29 normals and 10 out of 11 cases with frontal lobe injuries.

When, therefore, the patient is known on other grounds to be neither psychotic nor neurotic, this battery of tests offers a very accurate indication of whether or not the lesion is situated in the frontal lobes. The impairment index was validated on a group different from the standardization group and was repeated on an independent group. The only obvious objection to the index is the inadequate representation of groups other than normals or brain-damaged.

Perceptual and Motor Tests of Brain Damage

Many attempts have been made to use the Rorschach test for the diagnosis of brain damage. The approach has usually been made in terms of signs specific to brain damage. The approach is basically similar to that used in tests employing the concept of deterioration, for the signs are taken to indicate inadequate or lowered performance. One of the earliest approaches was made by Piotrowski (49, 50), who used 33 records, consisting of 18 brain-damaged cases, 10 cases with noncerebral disturbance of the central nervous system, and 5 cases of conversion hysteria. Ten signs were selected as differentiating between the three groups. These signs were very carefully defined by Piotrowski. The organic group produced a mean of 6.2 signs; the other groups, a mean of 1.5 signs; there was no overlap between the groups. Thus, it was considered that the presence of five or more of the Piotrowski signs was in-

dicative of brain damage. In a later paper (51), Piotrowski showed that the number of signs produced was a function, in part, of the severity of the personality changes produced by the disorder, and that the number of signs also increased with age. He also pointed out that some of his signs were produced by schizophrenics and neurotics. He claimed that, out of 56 patients producing five or more signs, 55 were in fact organic; of 25 patients originally considered, but later rejected as such by neurological tests, only one produced an organic record.

Ross (54) tried to repeat the findings of Piotrowski. He used two groups which closely approximated those used by Piotrowski. He also tested several other groups and found that the signs occurred highly significantly more often in the group with cerebral lesions, and significantly more often in the epileptics, than in all the other groups together. However, although five or more of these signs were found most often in patients with disease of the cerebral cortex and subcortical tissue, they were not specific for these lesions. Thus, 55 per cent of brain-damaged patients showed five or more signs, but so did 30 per cent with noncortical lesions of the central nervous system, and also 20 per cent of the psychotics and 14 per cent of the neurotics. Furthermore, Ross showed that 50 per cent of the cortical cases and 53 per cent of the epileptic cases showed five or more of the Harrower-Erickson signs of neurosis.

In two later papers (55, 57), Ross divided the signs into four groups: those common to neurotic and organic patients; the neurotic differential signs; the organic differential signs; and the organic excluding signs. Each sign was then weighted in rough proportion to its differential

incidence in the groups of individuals being compared. The four sets of scores were then combined for each individual to give two ratings. These two ratings were called the "instability" and "disability" ratings, and were standardized on 50 neurotics, 24 organics, 50 superior, and 50 average normals. This represented an advance on Piotrowski's method, in that a patient could be given a rating on both the organic and the neurotic dimensions.

The next important contribution came from Hughes (32, 33), who derived 14 signs from a factor analysis of the 22 signs he found in the literature on brain damage. Different weights were assigned to these signs according to their discriminatory power. He used 218 subjects, including 50 with brain damage, 68 schizophrenics, 74 neurotics, 4 manic-depressives, and 22 normals. The point-biserial correlation between presence or absence of brain damage and score on the sign pattern was +.79, which was highly significant. When a cutting score of seven or above was used, 82 per cent of the organics were correctly identified, while only 1 per cent of nonorganics were falsely identified as organic. Using Piotrowski's signs, Hughes could correctly identify only 20 per cent of the organics without including many nonorganics.

However, these encouraging results were shown to have a serious flaw by the recent work of Diers and Brown (16). They took a group of 25 multiple sclerotics and divided them into two groups, 15 with an IQ above 102, and 10 with an IQ below 102. This latter group was carefully matched with 11 nonorganic patients with IQ's below 102. The three groups were then tested and the results analyzed for Hughes's signs. The mean number of signs for the

more intelligent organics was only 0.33, none obtaining a score above seven, and ten obtaining a zero or negative score. As opposed to this the organics of average or dull IQ obtained a mean of 5.2 signs, and the nonorganics of average or dull IQ obtained a mean of 4.18 signs. None of the patients in these groups obtained a negative score. Diers and Brown concluded that there is an inverse relationship between IQ measurement on the Wechsler and the weighted Hughes score, independent of the factor of intracranial pathology. The Hughes signs are, therefore, invalid, unless obtained with a patient of high intelligence—and the study showed that, in fact, none of the intelligent organics used by Diers and Brown fell within the organic range.

A new approach was made to the problem by Dörken and Kral (17), who, instead of asking what signs the organic patient would show on the Rorschach, investigated what signs he would be likely *not* to show. Seven signs were determined, and each sign was weighted according to degree of occurrence in organic or nonorganic states. By this means, a total possible score of ten was determined. In terms of the frequency distribution, the cutting point was fixed between two and three; that is, scores from three to ten, inclusive, exclude a diagnosis of brain damage. Using this method, they claimed that 92.9 per cent of organics were identified, and 83.3 per cent of nonorganics were identified as such. If Piotrowski's signs were used however, only 50 per cent of the organics would be identified; while use of the Ross "disability" ratio identified 75 per cent of the organics.

One additional study may be mentioned here. Buhler, Buhler, and Lefever (8), using 30 normals, 70 neu-

rotics, 50 psychopaths, 27 schizophrenics, and 30 organics, developed a Basic Rorschach Score. They claimed that this score was capable of separating clinical groups in a statistically reliable manner. But they also stated that

the variability of scores within each clinical group indicates that the Basic Rorschach Score alone is not a sufficient basis for individual diagnosis. The placement of the individual case on a scale of adjustment or ego-integration does, however, appear to be a highly probable outcome of the use of the Basic Rorschach Score (8, p. 161).

In other words, the technique is designed, not to discriminate between organic and other groups as such, but rather to give an estimate of the degree of mental illness—a procedure which is based, apparently, on the theory that persons vary on a single dimension from normal through neurotic to psychotic, and which has been criticized by Eysenck (18, 19). A more recent monograph by Buhler, Lefever, Kallstedt, and Peak (9) confirmed the results obtained in the initial study.

The major criticisms to be made of the Rorschach as a test of brain damage seem to be that in all the scoring systems, except that of Dörken and Kral, the greatest weighting is given to those factors that are the most difficult to score and that depend, therefore, to the highest extent on the subjective evaluation of the examiner. Second, although all authors using the sign approach seem to obtain adequate differentiation of groups for their particular study, the differentiating power invariably drops considerably when these methods are repeated by other workers. Thus, where Piotrowski claimed powerful discrimination for his signs, Dörken and Kral found that they identified only 50 per cent of their

organic group; and other groups of signs used by Ross declined to an efficiency of 75 per cent. Third, most of the studies have ignored the influence of age and intelligence; the classic example of this is, of course, the work of Hughes, which at first sight seemed so promising. We must conclude, therefore, that, while most workers in this field report satisfactory discrimination, the constant "dog eat dog" method by which one set of signs is set aside as unusable and replaced by a new set by subsequent workers does not inspire confidence in the validity of the most recent method, that of Dörken and Kral. That the Rorschach offers distinct promise in this problem cannot be denied; that it has been shown to be a satisfactory test of brain damage is open to question.

Several tests have been recently constructed which are not very well known. Two of these are described in the monograph by Armitage (4). In their original form, these tests are known as the Trail Making Test and the Patch Test. The Trail Making Test consists of two parts. Part A consists of a sheet of paper on which there is printed a series of circles. In the center of each of these is a number, with a range for the test proper of 1–25, and for the sample, which is on the opposite side of the page, of 1–7. These circles are spatially arranged in a random order. The patient is required to draw a line between the circles, starting at number 1 and finishing at number 25. He is asked to work as rapidly as possible and to erase errors. Part B is very similar, but more complex, in that the numbered circles are mixed with circles lettered from A onwards. The task is to draw a line from circle one to circle A; then to go to circle two, then to circle B, and so on. Scoring

is in terms of time and accuracy. If no errors are made, or if an error is corrected very quickly, the score obtained depends directly on the speed with which the test is completed. If the item is accomplished in less than 20 seconds, a maximum score of ten is given; however, if longer times are taken, partial credits are given. The Patch Test requires the duplication of nine colored circular patterns, one serving as a demonstration design. For every pattern, the materials provided consist of 19 paper circles of different colors. Some of these are solid, while others have different sections of the interior part removed so that, by placing them on top of one another, a pattern may be formed. The test materials are arranged in front of the patient in a standard order. The eight test designs are in graded order of difficulty, and the test is terminated after two failures. Each design, with one exception, can be duplicated only by putting together specific pieces out of the 19 provided. In the standardization of these tests, Armitage used as subjects 44 patients known to have sustained brain damage (9 focal, 17 focal-diffuse, 18 diffuse). The control group consisted of 45 normal subjects and 16 mild neurotics. The groups were considered roughly comparable as to age, level of education, and pre-injury occupation.

It was found that Trail Making Test A, the simplest of the tests, discriminated the brain-damaged group best from the two control groups. With a cutting score of 10, the total misclassification was 16 out of 95, 32 of the 44 organics being positively identified. Trail Making Test B identified 39 out of 44 organics, but misclassified 15 out of 51 normals; and a combined score was not more effective. As a clinical instrument, the Patch Test identified 26 out of 43 organics, misclassifying 7 out of 51 controls. The tests would appear to be useful ones and merit further study. Unfortunately, no data are given on the performance of psychotic patients, and the study has not been repeated.

The genesis and construction of the Block Design Rotation Test has been described in three articles by Shapiro (61, 62, 63). This test resulted from the observation that some patients, while doing the Goldstein block design test, left the completed design in a rotated position compared with the test figure. This rotation might be as great as 45°, but rarely exceeded it. Various hypotheses were examined and it was found that the rotation effect could be maximized when three factors were interrelated in a certain way— namely, the factors of figure shape, ground shape, and angle of the line of symmetry (the line of symmetry being defined as the line which divided a design into mirrored halves). Thus, rotation was found to be maximal when the figure and ground shapes were in a diamond orientation and the line of symmetry was at an angle.

From various theoretical and empirical considerations, it was deduced that the rotation effect would be maximized under these conditions in brain-damaged patients. Accordingly, a special set of cards was devised, 40 in all. Each design could be reproduced by using four Kohs blocks. Two groups of subjects, 19 brain-damaged and 19 psychiatric patients (the latter without brain damage), were carefully matched for age and sex, and given the test under identical conditions. Results were according to prediction, the brain-

damaged patients rotating an average of 8° per card, while the functional patients rotated an average of only about 2° per card. Considering the matched pairs together, every brain-damaged patient rotated more than the corresponding functional patient, with one exception—and this patient's diagnosis was later changed independently to one of compensation neurosis. With a cutting score of 6° rotation per card, the test identified 14 out of the 19 brain-damaged patients, but misclassified only 1 functional patient. Nearly all the functional patients rotated a total of less than 200° for the 40 cards, and only one rotated more than 250°. Many brain-damaged patients, on the other hand, rotated more than 400° or 500° for the 40 cards, and much higher totals have been encountered in clinical practice.

When a further group of 19 brain-damaged patients (this time consisting of patients who had actually been operated) was tested, a closely similar pattern of results was found. Data on several normal control groups showed that the test also discriminated between normals and brain-damaged patients, and revealed the curious and unexpected fact that, as a group, the controls rotated significantly more than the functionals (though this was significant at the .05 level only). The effects of age, sex, and intelligence have also been calculated and will be published shortly. Thus, the rotation test satisfies the criteria laid down in this paper in that it provides data on at least three groups of subjects, it takes into account factors such as intelligence, age, and sex, and its results have subsequently been confirmed on new, independent groups. It is also an entirely objectively scored test, the amount of rotation being re-corded (in the later studies) photographically, with reference to a constant base line.

Using the original two groups of subjects, Shapiro (unpublished data) also found that the manual dexterity test and the finger dexterity test of the U. S. Employment Service battery of tests discriminated very significantly between brain-damaged and functional patients. The former test identified 14 out of the 19 brain-damaged patients and misclassified 3 out of the 19 functionals; the latter identified 12 out of the 19 brain-damaged patients and misclassified 5 out of the 19 functionals. Used together, the tests successfully identified 16 of the brain-damaged patients.

DISCUSSION

It is doubtful whether any aspect of psychological testing has been more inadequately treated than the diagnostic assessment of brain damage. From a wide range of possible criticisms, some of the most obvious will be cited:

a. In many instances, validation studies have not used comparable groups. For example, the majority of cases in Hunt's (34) brain-damaged group were paretics and others suffering from diffuse brain damage. In his validation study, Armitage (4), on the other hand, included many patients suffering from traumatic head injury as the result of penetrating missiles. There is no reason to suppose that these two groups were at all comparable with regard to type, locus, or severity of brain damage. Armitage's procedure may be justified, however, in so far as Hunt described his tests as a measure of brain damage without further qualification.

b. Few authors, when constructing their tests, choose their cases with sufficient care, in most instances considering brain damage as a unitary factor. That this is unsound from the point of view of diagnosis may be shown by evidence from many sources. Thus, if a random group of brain-damaged patients is given a test, the results nearly always fall into an abnormal distribution (i.e., are skewed). Many of the brain-damaged group

behave on a given test like normal controls or functional patients, while others obtain very abnormal scores. Frequently, this sort of pattern accounts for the significance of differences between the groups. This kind of distribution was encountered in the Block Design Rotation Test and was recently seen in a study by Battersby, Teuber, and Bender (6) on the behavior of brain-damaged patients in a problem-solving situation. From an anatomical and physiological standpoint, there is likewise no reason why all brain-damaged patients should belong together. As Penfield and Evans (48) pointed out, there is a wealth of difference between the brain damage resulting from scar formation on the temporal lobe following an accident, and the scar formation resulting from a temporal lobectomy. Similarly, there is no reason to suppose that a leucotomy operation will have the same effect as diffuse brain damage.

c. Many investigators neglect almost completely the elementary necessity for evaluating and controlling various relevant factors such as age, sex, and intelligence. The Shipley-Hartford Retreat Scale, the Rorschach, and the Goldstein test are noteworthy examples where failure to do this has led to ambiguity in the clinical use of the test. That factors such as these cannot be neglected without serious risk of error is apparent from an article by Hebb (30). Using simple patterns that had to be reproduced with pieces of wood, Hebb found that "no pattern could be devised, which was so easy that all patients in the public wards of a general hospital could succeed with it in one minute, even though other tests showed that one was not dealing with a population of mental defectives" (30, p. 16). Hebb concludes that, although this kind of material tends to be eliminated in tests which are adequately standardized, "in special tests which have not been standardized, there is a real danger of assuming that a variation from the norm, which is frequently obtained for the normal population, can be due only to the effects of cerebral injury" (30, p. 17). In addition, many authors fail to state whether or not their subjects were tested for specific defects, even when this is clearly necessary. The Trail Making and Patch tests, for instance, might well be performed badly by normal persons with defective eyesight. These tests might be peculiarly successful in identifying soldiers suffering from penetrating wounds as brain damaged and be completely unsuccessful in cases of diffuse brain damage.

d. Many tests employ very unreliable and subjective scoring systems. This applies particularly to the Rorschach, where most weight

is given to factors such as impotence, perplexity, etc., but also to the Hunt-Minnesota Test, where, for example, more credit is given for a response within one-half of a second than for a response within one-half to one second.

e. Occasionally, there is failure to realize that a test may discriminate between two groups statistically at a high level of significance, and yet still be unusable clinically because the misclassification would be very high.

f. In many tests, there is failure to control relevant variables. Thus, in the Hunt-Minnesota test, it is claimed that immediate and long-term memory are being measured. Examination of the items, however, reveals failure to control for differential learning ability.

These criticisms are made on methodological grounds; there seems no reason why they should not be overcome. Assuming that they were overcome, would it be possible to develop adequate tests of brain damage? The answer would appear to be in the negative, *as long as a purely engineering approach is made.* Thus a most painstaking study by Lynn, Levine, and Hewson (42) was concerned with the aftereffects of exposure to blast during the war. Such exposure might result in a closed head injury with accompanying residual symptomatology determined by the brain damage alone; or in a neurotic syndrome, identical clinically to that attributable to brain damage, but without any actual trauma to the brain; or (most commonly) in a combination of the two. Starting with over 4,000 patients, they reduced this group to 81 "pure" cases and attempted to differentiate these by means of a battery of tests. Elaborate precautions were taken to minimize unreliability of diagnosis and to control for age, sex, intelligence, etc., and a large number of independent validating groups were used. Nevertheless, for these "pure" groups the misclassification was still 12 per cent. For the validation groups, it is

arguable that the statistical techniques were inadequate, but the results are still disappointing.

The basic flaw in most of the tests discussed above lies, however, in the theoretical approach. Examination of the tests suggests that one major hypothesis concerning the psychological effects of brain damage permeates the work; this is, that brain damage results in deterioration of a relatively permanent nature. Such a theory is unlikely to provide a satisfactory basis for the construction of tests of brain damage because it is not exclusive to brain damage. This criticism seems more important than the usual one that deterioration is an ill-defined concept—so much so that Hunt (35) preferred the neutral term "psychological deficit." Similarly, provided the test is rigorously scored, the assumption that vocabulary level is resistant to the effects of mental illness has been shown to be false by the work of Yacorzynski (69), Capps (12), and Simmins (66); while Crown (14), using 4 cases of myxedema (which causes apparent intellectual deterioration) tested before and after treatment, found that the Shipley vocabulary MA of these patients rose by an average of 10.2 months following treatment by thyroxine, and on another vocabulary test, the mean vocabulary IQ rose by 6 and 14 points in two of the cases.

The first essential in the construction of tests of brain damage, therefore, is the development of a theory which is exclusive to brain damage. One possible theory has been presented by Shapiro (61, 62, 63) in his articles, and may be stated as follows: *Brain damage results in the development of states of exaggerated inhibition*. Such a theory is more satisfactory than a theory of deterioration (in so far as it is supported by experimental evidence) because it would not be postulated to account for functional disorder or for the behavior of normals. Hence, it becomes possible to make deductions and set up experimental situations in which brain-damaged patients behave differently from others, and through which verification of the theory and development of tests of brain damage are feasible. Having done this and verified the theory in its broad outlines, it now becomes possible to consider alternative hypotheses and also to consider why some brain-damaged patients do not behave as predicted. One immediate possibility, of course, is location of the damage. Halstead (28) and Rylander (58, 59) concur in demonstrating that it appears to be much easier to distinguish brain damage in the frontal lobes from that in other areas of the brain. Another possibility is the significance of previous personality. The interaction of personality structure and the effects of brain damage have been largely ignored by most investigators, although it is known that extensive brain damage may have little or no effect on functioning, and that a brain-damaged patient may be neurotic or psychotic. Now Eysenck (18) has shown that there are tests which discriminate between psychotics and normals but not between neurotics and normals, and vice versa; he has argued, on this basis, that neuroticism and psychoticism are orthogonal (independent) factors. At this stage, therefore, it would be possible to construct a battery of tests which would give a particular patient a factor score on each of these three dimensions —psychoticism, neuroticism, and brain damage. The question to be answered would then become, not, is the patient psychotic *or* neurotic *or* brain

...maged, but what is his perform-...e with respect to these three vari-...bs, and what is their interrelation-...p. At this point, the internal valid-...n of the tests would become im-...ortant since it would be necessary ...o show that those tests measuring ...the effects of brain damage have sig-...nificantly positive intercorrelations ...among themselves, but not with the ...ests of psychoticism or neuroticism. It is clear, therefore, that the prob-lem is much more complex than was suggested in the early part of this paper. In the view of the writer, a purely empirical approach is unlikely to yield satisfactory results, nor is an approach based on a theory which has not been adequately tested ex-perimentally. A satisfactory test of brain damage should be based on a reasonable theory that has been ex-perimentally tested, has been sup-ported by adequate statistical treat-ment, and has taken into account all relevant variables. Such an ap-proach would at least help to over-come the impasse which seems to have been reached with many of the tests reviewed above.

REFERENCES

1. AITA, J. A., ARMITAGE, S. G., REITAN, R. M., & RABINOWITZ, A. The use of certain psychological tests in the evalua-tion of brain injury. *J. gen. Psychol.*, 1947, 37, 25–44.

2. ALLEN, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, 3, 225–230.

3. ANDERSEN, A. L. The effect of laterality of localization of brain damage on Wechs-ler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1950, 6, 191–194.

4. ARMITAGE, S. G. An analysis of certain psychological tests used for the evalua-tion of brain injury. *Psychol. Monogr.*, 1946, 60, No. 1 (Whole No. 277).

5. ASH, P. The reliability of psychiatric diagnoses. *J. abnorm. soc. Psychol.*, 1949, 44, 272–276.

6. BATTERSBY, W. S., TEUBER, H. L., & BENDER, M. B. Problem solving be-havior in men with frontal or occipital brain injuries. *J. Psychol.*, 1953, 35, 329–351.

7. BOYD, H. F. A provisional quantitative scoring with preliminary norms for the Goldstein-Scheerer Cube Test. *J. clin. Psychol.*, 1949, 5, 148–153.

8. BUHLER, CHARLOTTE, BUHLER, K., & LEFEVER, D. W. *Development of the Basic Rorschach Score with manual of directions.* Los Angeles, Calif.: Ror-schach Standardization Studies, No. I, 1948.

9. BUHLER, CHARLOTTE, LEFEVER, D. W., KALLSTEDT, FRANCES E., & PEAK, H. M. *Development of the Basic Ror-schach Score. Supplementary monograph.*

Los Angeles, Calif.: Rorschach Stand-ardization Studies, 1952.

10. BUROS, O. K. (Ed.) *The third mental meas-urements yearbook.* New Brunswick: Rutgers Univer. Press, 1949.

11. CANTER, A. H. Direct and indirect meas-ures of psychological deficit in multiple sclerosis. Part II. *J. gen. Psychol.*, 1951, 44, 27–50.

12. CAPPS, H. M. Vocabulary changes in mental deterioration. *Arch. Psychol.*, 1939, 34, No. 242.

13. CORSINI, R. J., & FASSETT, KATHERINE K. The validity of Wechsler's mental de-terioration index. *J. consult. Psychol.*, 1952, 16, 462–468.

14. CROWN, S. Notes on an experimental study of intellectual deterioration. *Brit. med. J.*, 1949, 2, 684–685.

15. DIERS, W. C., & BROWN, C. C. Psycho-metric patterns associated with multi-ple sclerosis. I. Wechsler-Bellevue pat-terns. *Arch. Neurol. Psychiat.*, 1950, 63, 760–765.

16. DIERS, W. C., & BROWN, C. C. Rorschach "organic signs" and intelligence level. *J. consult. Psychol.*, 1951, 15, 343–345.

17. DÖRKEN, H., & KRAL, V. A. The psycho-logical differentiation of organic brain lesions and their localization by means of the Rorschach test. *Amer. J. Psy-chiat.*, 1952, 108, 764–771.

18. EYSENCK, H. J. *The scientific study of per-sonality.* London: Routledge & Kegan Paul, 1952.

19. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.

20. FLEMING, G. W. T. H. The Shipley-

Hartford Retreat Scale for measuring intellectual impairment. A preliminary communication. *J. ment. Sci.*, 1943, 89, 64–68.

21. GARFIELD, S. L., & FEY, W. F. A comparison of the Wechsler-Bellevue and Shipley-Hartford scales as measures of mental impairment. *J. consult. Psychol.*, 1948, 12, 259–264.

22. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).

23. GRASSI, J. R. The Fairfield Block Substitution Test for measuring intellectual impairment. *Psychiat. Quart.*, 1947, 21, 474–489.

24. GRASSI, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology.* Springfield, Ill.: Charles C Thomas, 1953.

25. GUTMAN, BRIGETTE. The application of the Wechsler-Bellevue scale in the diagnosis of organic brain disorders. *J. clin. Psychol.*, 1950, 6, 195–198.

26. HALL, K. R. L. The testing of abstraction with special reference to impairment in schizophrenia. *Brit. J. med. Psychol.*, 1951, 24, 118–131.

27. HALSTEAD, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and nonequivalent stimuli. *Amer. J. Psychiat.*, 1940, 96, 1263–1294.

28. HALSTEAD, W. C. *Brain and intelligence. A quantitative study of the frontal lobes.* Chicago: Univer. of Chicago Press, 1947.

29. HALSTEAD, W. C., & SETTLAGE, P. H. Grouping behavior of normal persons and of persons with lesions of the brain. Further analysis. *Arch. Neurol. Psychiat., Chicago*, 1943, 49, 489–506.

30. HEBB, D. O. Man's frontal lobes. *Arch. Neurol. Psychiat., Chicago*, 1945, 54, 10–24.

31. HEWSON, LOUISE, R. The Wechsler-Bellevue scale and the substitution test as aids in neuropsychiatric diagnosis. *J. nerv. ment. Dis.*, 1949, 109, 158–183, 246–266.

32. HUGHES, R. M. Rorschach signs for the diagnosis of organic pathology. *J. proj. Tech.*, 1948, 12, 165–167.

33. HUGHES, R. M. A factor analysis of Rorschach diagnostic signs. *J. gen. Psychol.*, 1950, 43, 85–103.

34. HUNT, H. F. A practical clinical test for organic brain damage. *J. appl. Psychol.*, 1943, 27, 375–386

35. HUNT, J. McV. (Ed.) *Personality and behavior disorders.* Vol. 2. New York: Ronald, 1944.

36. HUTTON, E. L. The investigation of personality in patients treated by prefrontal leucotomy. *J. ment. Sci.*, 1942, 88, 275–281.

37. JUCKEM, HARRIET, J., & WOLD, JANET. A study of the Hunt-Minnesota test for organic brain damage at the upper levels of vocabulary. *J. consult. Psychol.*, 1948, 12, 53–57.

38. KASS, W. Wechsler's mental deterioration index in the diagnosis of organic brain disease. *Trans. Kans. Acad. Sci.*, 1949, 52, 66–70.

39. KOBLER, F. J. The measurement of improvement among neuropsychiatric patients in an army convalescent center. *J. clin. Psychol.*, 1947, 3, 121–128.

40. LEVI, J., OPPENHEIM, SADIE, &. WECHSLER, D. Clinical use of the mental deterioration index of the Bellevue-Wechsler scale. *J. abnorm. soc. Psychol.*, 1945, 40, 405–407.

41. LIDZ, T., GAY, J. R., & TIETZE, C. Intelligence in cerebral deficit states and schizophrenia measured by Kohs block test. *Arch. Neurol. Psychiat., Chicago*, 1942, 48, 568–582.

42. LYNN, J. G., LEVINE, KATE N., & HEWSON, LOUISE R. Psychologic tests for the clinical evaluation of late "diffuse organic," "neurotic," and "normal" reactions after closed head injury. *Proc. Ass. Res. nerv. ment. Dis.*, 1945, 24, 296–378.

43. McFIE, J., & PIERCY, M. F. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, 75, 292–311.

44. MAGARET, ANN, & SIMPSON, MARY M. A comparison of two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, 12, 265–269.

45. MALAMUD, RACHEL F. Validity of the Hunt-Minnesota Test for Organic Brain Damage. *J. appl. Psychol.*, 1946, 30, 271–275.

46. MANSON, M. P., & GRAYSON, H. M. The Shipley-Hartford Retreat Scale as a measure of intellectual impairment for military prisoners. *J. appl. Psychol.*, 1947, 31, 67–81.

47. MEEHL, P. E., & JEFFERY, MARY. The Hunt-Minnesota Test for Organic Brain Damage in cases of functional depression. *J. appl. Psychol.*, 1946, 30, 276–287.

48. PENFIELD, W., & EVANS, J. The frontal lobe in man: a clinical study of maximum removals. *Brain*, 1935, **58**, 115–133.

49. PIOTROWSKI, Z. The Rorschach inkblot method in organic disturbance of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.

50. PIOTROWSKI, Z. Rorschach studies of cases with lesions of the frontal lobes. *Brit. J. med. Psychol.*, 1938, **17**, 105–118.

51. PIOTROWSKI, Z. Positive and negative Rorschach organic reactions. *Rorschach Res. Exch.*, 1940, **4**, 147–151.

52. REYNELL, W. R. A psychometric method of determining intellectual loss following head injury. *J. ment. Sci.*, 1944, **90**, 710–719.

53. ROGERS, L. S. A comparative evaluation of the Wechsler-Bellevue mental deterioration index for various adult groups. *J. clin. Psychol.*, 1950, **6**, 199–202.

54. ROSS, W. D. The contribution of the Rorschach method to clinical diagnosis. *J. ment. Sci.*, 1941, **87**, 331–348.

55. ROSS, W. D. A quantitative use of the Rorschach method. "Instability" and "disability" ratings which show clinical and psychometric correlations. *Amer. J. Psychiat.*, 1944, **101**, 100–104.

56. ROSS, W. D., & MCNAUGHTON, F. L. Head injury: a study of patients with chronic posttraumatic complaints. *Arch. Neurol. Psychiat., Chicago*, 1944, **52**, 255–269.

57. ROSS, W. D., & ROSS, SALLY. Some Rorschach ratings of clinical value. *Rorschach Res. Exch.*, 1944, **8**, 1–9.

58. RYLANDER, G. *Personality changes after operations on the frontal lobes.* London: Oxford Univer. Press, 1939.

59. RYLANDER, G. Mental changes after excision of cerebral tissue. A clinical study of 16 cases of resections in the parietal, temporal and occipital lobes. *Acta Psychiat. Neurol.*, 1943, Suppl. 25.

60. SCHERRER, M. An experiment in abstraction. Testing form-disparity tolerance. *Conf. near J.* 1949 9 sol 464

61. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. I. Initial experiments. *J. ment. Sci.*, 1951, 97, 90–110

62. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. II. Confirmatory and explanatory experiments. *J. ment. Sci.*, 1952, **98**, 605–17.

63. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, 99, 394–409.

64. SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, **9**, 371–377.

65. SHIPLEY, W. C., & BURLINGAME, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, **97**, 1313–1325.

66. SIMMINS, C. A. Studies in experimental psychiatry: IV. Deterioration of g in psychotic patients. *J. ment. Sci.*, 1933, **79**, 704–734.

67. TOOTH, G. On the use of mental tests for the measurement of disability after head injury. With a comparison between the results of these tests in patients after head injury and psychoneurotics. *J. Neurol. Neurosurg. Psychiat.*, 1947, **10**, 1–11.

68. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

69. YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock deterioration test. *Psychol. Rev.*, 1941, **48**, 261–267.

THE THEORY OF DECISION MAKING[1]

WARD EDWARDS
The Johns Hopkins University

Many social scientists other than psychologists try to account for the behavior of individuals. Economists and a few psychologists have produced a large body of theory and a few experiments that deal with individual decision making. The kind of decision making with which this body of theory deals is as follows: given two states, A and B, into either one of which an individual may put himself, the individual chooses A in preference to B (or vice versa). For instance, a child standing in front of a candy counter may be considering two states. In state A the child has $0.25 and no candy. In state B the child has $0.15 and a ten-cent candy bar. The economic theory of decision making is a theory about how to predict such decisions.

Economic theorists have been concerned with this problem since the days of Jeremy Bentham (1748–1832). In recent years the development of the economic theory of consumer's decision making (or, as the economists call it, the theory of consumer's choice) has become exceedingly elaborate, mathematical, and voluminous. This literature is almost unknown to psychologists, in spite of sporadic pleas in both psychological (40, 84, 103, 104) and economic (101, 102, 123, 128, 199, 202) literature for greater communication between the disciplines.

The purpose of this paper is to review this theoretical literature, and also the rapidly increasing number of psychological experiments (performed by both psychologists and economists) that are relevant to it. The review will be divided into five sections: the theory of riskless choices, the application of the theory of riskless choices to welfare economics, the theory of risky choices, transitivity in decision making, and the theory of games and of statistical decision functions. Since this literature is unfamiliar and relatively inaccessible to most psychologists, and since I could not find any thorough bibliography on the theory of choice in the economic literature, this paper includes a rather extensive bibliography of the literature since 1930.

THE THEORY OF RISKLESS CHOICES[2]

Economic man. The method of those theorists who have been con-

[2] No complete review of this literature is available. Kauder (105, 106) has reviewed the very early history of utility theory. Stigler (180) and Viner (194) have reviewed the literature up to approximately 1930. Samuelson's book (164) contains an illuminating mathematical exposition of some of the content of this theory. Allen (6) explains the concept of indifference curves. Schultz (172) re-

cerned with the theory of decision making is essentially an armchair method. They make assumptions, ...from those assumptions they de... ...the ones which presumably can be tested, though it sometimes seems unlikely that the testing will ever occur. The most important set of assumptions made in the theory of riskless choices may be summarized by saying that it is assumed that the person who makes any decision to which the theory is applied is an economic man.

What is an economic man like? He has three properties. (a) He is completely informed. (b) He is infinitely sensitive. (c) He is rational.

Complete information. Economic man is assumed to know not only what all the courses of action open to him are, but also what the outcome of any action will be. Later on, in the sections on the theory of risky choices and on the theory of games, this assumption will be relaxed somewhat. (For the results of attempts to introduce the possibility of learning into this picture, see 51, 77.)

Infinite sensitivity. In most of the older work on choice, it is assumed that the alternatives available to an individual are continuous, infinitely divisible functions, that prices are infinitely divisible, and that economic man is infinitely sensitive. The only purpose of these assumptions is to make the functions that they lead to,

continuous and differentiable. Some 182 has recently shown that they can be abandoned with no serious changes in the theory of choice.

Rationality. The crucial fact about economic man is that he is rational. This means two things: He can weakly order the states into which he can get, and he makes his choices so as to maximize something.

Two things are required in order for economic man to be able to put all available states into a weak ordering. First, given any two states into which he can get, A and B, he must always be able to tell either that he prefers A to B, or that he prefers B to A, or that he is indifferent between them. If preference is operationally defined as choice, then it seems unthinkable that this requirement can ever be empirically violated. The second requirement for weak ordering, a more severe one, is that all preferences must be transitive. If economic man prefers A to B and B to C, then he prefers A to C. Similarly, if he is indifferent between A and B and between B and C, then he is indifferent between A and C. It is not obvious that transitivity will always hold for human choices, and experiments designed to find out whether or not it does will be described in the section on testing transitivity.

The second requirement of rationality, and in some ways the more important one, is that economic man must make his choices in such a way as to maximize something. This is the central principle of the theory of choice. In the theory of riskless choices, economic man has usually been assumed to maximize utility. In the theory of risky choices, he is assumed to maximize expected utility. In the literature on statistical decision making and the theory of games, various other fundamental

---

views the developments up to but not including the Hicks-Allen revolution from the point of view of demand theory. Hicks's book (87) is a complete and detailed exposition of most of the mathematical and economic content of the theory up to 1939. Samuelson (167) has reviewed the integrability problem and the revealed preference approach. And Wold (204, 205, 206) has summed up the mathematical content of the whole field for anyone who is comfortably at home with axiom systems and differential equations.

principles of decision making are considered, but they are all maximization principles of one sort or another.

The fundamental content of the notion of maximization is that economic man always chooses the best alternative from among those open to him, as he sees it. In more technical language, the fact that economic man prefers A to B implies and is implied by the fact that A is higher than B in the weakly ordered set mentioned above. (Some theories introduce probabilities into the above statement, so that if A is higher than B in the weak ordering, then economic man is more likely to choose A than B, but not certain to choose A.)

This notion of maximization is mathematically useful, since it makes it possible for a theory to specify a unique point or a unique subset of points among those available to the decider. It seems to me psychologically unobjectionable. So many different kinds of functions can be maximized that almost any point actually available in an experimental situation can be regarded as a maximum of some sort. Assumptions about maximization only become specific, and therefore possibly wrong, when they specify what is being maximized.

There has, incidentally, been almost no discussion of the possibility that the two parts of the concept of rationality might conflict. It is conceivable, for example, that it might be costly in effort (and therefore in negative utility) to maintain a weakly ordered preference field. Under such circumstances, would it be "rational" to have such a field?

It is easy for a psychologist to point out that an economic man who has the properties discussed above is very unlike a real man. In fact, it is so easy to point this out that psycholo-

gists have tended to reject out of hand the theories that result from these assumptions. This isn't fair. Surely the assumptions contained in Hullian behavior theory (91) or in the Estes (60) or Bush-Mosteller (36, 37) learning theories are no more realistic than these. The most useful thing to do with a theory is not to criticize its assumptions but rather to test its theorems. If the theorems fit the data, then the theory has at least heuristic merit. Of course, one trivial theorem deducible from the assumptions embodied in the concept of economic man is that in any specific case of choice these assumptions will be satisfied. For instance, if economic man is a model for real men, then real men should always exhibit transitivity of real choices. Transitivity is an assumption, but it is directly testable. So are the other properties of economic man as a model for real men.

Economists themselves are somewhat distrustful of economic man (119, 156), and we will see in subsequent sections the results of a number of attempts to relax these assumptions.

Early utility maximization theory. The school of philosopher-economists started by Jeremy Bentham and popularized by James Mill and others held that the goal of human action is to seek pleasure and avoid pain. Every object or action may be considered from the point of view of pleasure- or pain-giving properties. These properties are called the *utility* of the object, and pleasure is given by positive utility and pain by negative utility. The goal of action, then, is to seek the maximum utility. This simple hedonism of the future is easily translated into a theory of choice. People choose the alternative, from among those open to them, that

leads to the greatest excess of positive over negative utility. This notion of utility maximization is the essence of the utility theory of choice. It will reappear in various forms throughout this paper. (Bohnert [30] discusses the logical structure of the utility concept.)

This theory of choice was embodied in the formal economic analyses of all the early great names in economics. In the hands of Jevons, Walras, and Menger it reached increasingly sophisticated mathematical expression and it was embodied in the thinking of Marshall, who published the first edition of his great *Principles of Economics* in 1890, and revised it at intervals for more than 30 years thereafter (137).

The use to which utility theory was put by these theorists was to establish the nature of the demand for various goods. On the assumption that the utility of any good is a monotonically increasing negatively accelerated function of the amount of that good, it is easy to show that the amounts of most goods which a consumer will buy are decreasing functions of price, functions which are precisely specified once the shapes of the utility curves are known. This is the result the economists needed and is, of course, a testable theorem. (For more on this, see **87, 159.**)

Complexities arise in this theory when the relations between the utilities of different goods are considered. Jevons, Walras, Menger, and even Marshall had assumed that the utilities of different commodities can be combined into a total utility by simple addition; this amounts to assuming that the utilities of different goods are independent (in spite of the fact that Marshall elsewhere discussed the notions of competing goods, like soap and detergents, and completing goods, like right and left shoes, which obviously do not have independent utilities. Edgeworth (53), who was concerned with such nonindependent utilities, pointed out that total utility was not necessarily an additive function of the utilities attributable to separate commodities. In the process he introduced the notion of indifference curves, and thus began the gradual destruction of the classical utility theory. We shall return to this point shortly.

Although the forces of parsimony have gradually resulted in the elimination of the classical concept of utility from the economic theory of riskless choices, there have been a few attempts to use essentially the classical theory in an empirical way. Fisher (63) and Frisch (75) have developed methods of measuring marginal utility (the change in utility $[u]$ with an infinitesimal change in amount possessed $[Q]$, i.e., du/dQ) from market data, by making assumptions about the interpersonal similarity of consumer tastes. Recently Morgan (141) has used several variants of these techniques, has discussed mathematical and logical flaws in them, and has concluded on the basis of his empirical results that the techniques require too unrealistic assumptions to be workable. The crux of the problem is that, for these techniques to be useful, the commodities used must be independent (rather than competing or completing), and the broad commodity classifications necessary for adequate market data are not independent. Samuelson (164) has shown that the assumption of independent utilities, while it does guarantee interval scale utility measures, puts unwarrantably severe restrictions on the nature of the resulting demand function. Elsewhere Samuelson (158) presented,

primarily as a logical and mathematical exercise, a method of measuring marginal utility by assuming some time-discount function. Since no reasonable grounds can be found for assuming one such function rather than another, this procedure holds no promise of empirical success. Marshall suggested (in his notion of "consumer's surplus") a method of utility measurement that turns out to be dependent on the assumption of constant marginal utility of money, and which is therefore quite unworkable. Marshall's prestige led to extensive discussion and debunking of this notion (e.g., 28), but little positive comes out of this literature. Thurstone (186) is currently attempting to determine utility functions for commodities experimentally, but has reported no results as yet.

Indifference curves. Edgeworth's introduction of the notion of indifference curves to deal with the utilities of nonindependent goods was mentioned above. An indifference curve is, in Edgeworth's formulation, a constant-utility curve. Suppose that we consider apples and bananas, and suppose that you get

FIG. 1. A HYPOTHETICAL INDIFFERENCE MAP

the same amount of utility from 10-apples-and-1-banana as you get from 6-apples-and-4-bananas. Then these are two points on an indifference curve, and of course there are an infinite number of other points on the same curve. Naturally, this is not the only indifference curve you may have between apples and bananas. It may also be true that you are indifferent between 13-apples-and-5-bananas and 5-apples-and-15-bananas. These are two points on another, higher indifference curve. A whole family of such curves is called an indifference map. Figure 1 presents such a map. One particularly useful kind of indifference map has amounts of a commodity on one axis and amounts of money on the other. Money is a commodity, too.

The notion of an indifference map can be derived, as Edgeworth derived it, from the notion of measurable utility. But it does not have to be. Pareto (146, see also 151) was seriously concerned about the assumption that utility was measurable up to a linear transformation. He felt that people could tell whether they preferred to be in state A or state B, but could not tell how much they preferred one state over the other. In other words, he hypothesized a utility function measurable only on an ordinal scale. Let us follow the usual economic language, and call utility measured on an ordinal scale *ordinal* utility, and utility measured on an interval scale, *cardinal* utility. It is meaningless to speak of the slope, or marginal utility, of an ordinal utility function; such a function cannot be differentiated. However, Pareto saw that the same conclusions which had been drawn from marginal utilities could be drawn from indifference curves. An indifference map can be drawn simply by finding all the com-
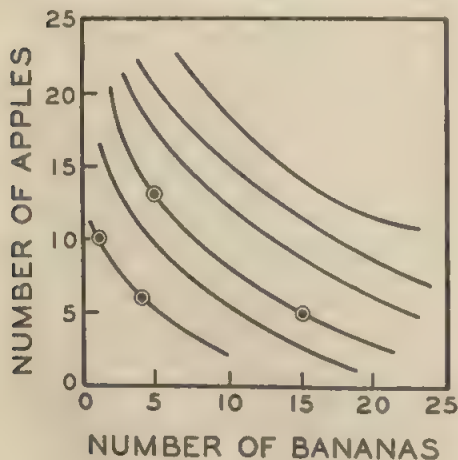
binations of the goods involved among which the person is indifferent. Pareto's formulation assumes that higher indifference curves have greater utility, but does not need to specify how much greater that utility is.

It turns out to be possible to deduce from indifference curves all of the theorems that were originally deduced from cardinal utility measures. This banishing of cardinal utility was furthered considerably by splendid mathematical papers by Johnson (97) and Slutsky (177). (In modern economic theory, it is customary to think of an n-dimensional commodity space, and of indifference hyperplanes in that space, each such hyperplane having, of course, $n-1$ dimensions. In order to avoid unsatisfactory preference structures, it is necessary to assume that consumers always have a complete weak ordering for all commodity bundles, or points in commodity space. Georgescu-Roegen [76], Wold [204, 205, 206, 208], Houthakker [90], and Samuelson [167] have discussed this problem.)

Pareto was not entirely consistent in his discussion of ordinal utility. Although he abandoned the assumption that its exact value could be known, he continued to talk about the sign of the marginal utility coefficient, which assumed that some knowledge about the utility function other than purely ordinal knowledge was available. He also committed other inconsistencies. So Hicks and Allen (88), in 1934, were led to their classic paper in which they attempted to purge the theory of choice of its last introspective elements. They adopted the conventional economic view about indifference curves as determined from a sort of imaginary questionnaire, and proceeded to derive all of the usual conclusions about

consumer demand with no reference to the notion of even ordinal utility (though of course the notion of an ordinal scale of preferences was still embodied in their derivation of indifference curves). This paper was for economics something like the behaviorist revolution in psychology.

Lange (116), stimulated by Hicks and Allen, pointed out another inconsistency in Pareto. Pareto had assumed that if a person considered four states, A, B, C, and D, he could judge whether the difference between the utilities of A and B was greater than, equal to, or less than the difference between the utilities of C and D. Lange pointed out that if such a comparison was possible for any A, B, C, and D, then utility was cardinally measurable. Since it seems introspectively obvious that such comparisons can be made, this paper provoked a flood of protest and comment (7, 22, 117, 147, 209). Nevertheless, in spite of all the comment, and even in spite of skepticism by a distinguished economist as late as 1953 (153), Lange is surely right. Psychologists should know this at once; such comparisons are the basis of the psychophysical Method of Equal Sense Distances, from which an interval scale is derived. (Samuelson [162] has pointed out a very interesting qualification. Not only must such judgments of difference be possible, but they must also be transitive in order to define an interval scale.) But since such judgments of differences did not seem to be necessary for the development of consumer demand theory, Lange's paper did not force the reinstatement of cardinal utility.

Indeed, the pendulum swung further in the behavioristic direction. Samuelson developed a new analytic foundation for the theory of con-

sumer behavior, the essence of which is that indifference curves and hence the entire structure of the theory of consumer choice can be derived simply from observation of choices among alternative groups of purchases available to a consumer (160, 161). This approach has been extensively developed by Samuelson (164, 165, 167, 169) and others (50, 90, 125, 126). The essence of the idea is that each choice defines a point and a slope in commodity space. Mathematical approximation methods make it possible to combine a whole family of such slopes into an indifference hyperplane. A family of such hyperplanes forms an indifference "map."

In a distinguished but inaccessible series of articles, Wold (204, 205, 206; see also 208 for a summary presentation) has presented the mathematical content of the Pareto, Hicks and Allen, and revealed preference (Samuelson) approaches, as well as Cassel's demand function approach, and has shown that if the assumption about complete weak ordering of bundles of commodities which was discussed above is made, then all these approaches are mathematically equivalent.

Nostalgia for cardinal utility. The crucial reason for abandoning cardinal utility was the argument of the ordinalists that indifference curve analysis in its various forms could do everything that cardinal utility could do, with fewer assumptions. So far as the theory of riskless choice is concerned, this is so. But this is only an argument for parsimony, and parsimony is not always welcome. There was a series of people who, for one reason or another, wanted to reinstate cardinal utility, or at least marginal utility. There were several mathematically invalid attempts to

show that marginal utility could be defined even in an ordinal-utility universe (23, 24, 163; 25, 114 Knight (110), in 1944, argued extensively for cardinal utility; he based his arguments in part on introspective considerations and in part on an examination of psychophysical scaling procedures. He stimulated a number of replies (29, 42; 111). Recently Robertson (154) pleaded for the reinstatement of cardinal utility in the interests of welfare economics (this point will be discussed again below). But in general the indifference curve approach, in its various forms, has firmly established itself as the structure of the theory of riskless choice.

Experiments on indifference curves. Attempts to measure marginal utility from market data were discussed above. There have been three experimental attempts to measure indifference curves. Schultz, who pioneered in deriving statistical demand curves, interested his colleague at the University of Chicago, the psychologist Thurstone, in the problem of indifference curves. Thurstone (185) performed a very simple experiment. He gave one subject a series of combinations of hats and overcoats, and required the subject to judge whether he preferred each combination to a standard. For instance, the subject judged whether he preferred eight hats and eight overcoats to fifteen hats and three overcoats. The same procedure was repeated for hats and shoes, and for shoes and overcoats. The data were fitted with indifference curves derived from the assumptions that utility curves fitted Fechner's Law and that the utilities of the various objects were independent. Thurstone says that Fechner's Law fitted the data better than the other possible functions he considered, but

presents no evidence for this assertion. The crux of the experiment was the attempt to predict the indifference curves between shoes and overcoats from the other indifference curves. This was done by using the other two indifference curves to infer utility functions for shoes and for overcoats separately, and then using these two utility functions to predict the total utility of various amounts of shoes and overcoats jointly. The prediction worked rather well. The judgments of the one subject used are extraordinarily orderly; there is very little of the inconsistency and variability that others working in this area have found. Thurstone says, "The subject . . . was entirely naive as regards the psychophysical problem involved and had no knowledge whatever of the nature of the curves that we expected to find" (185, p. 154). He adds, "I selected as subject a research assistant in my laboratory who knew nothing about psychophysics. Her work was largely clerical in nature. She had a very even disposition, and I instructed her to take an even motivational attitude on the successive occasions . . . I was surprised at the consistency of the judgments that I obtained, but I am pretty sure that they were the result of careful instruction to assume a uniform motivational attitude."[3] From the economist's point of view, the main criticism of this experiment is that it involved imaginary rather than real transactions (200).

The second experimental measurement of indifference curves is reported by the economists Rousseas and Hart (157). They required large numbers of students to rank sets of three combinations of different amounts of bacon and eggs. By assuming that all students had the same indifference curves, they were able to derive a composite indifference map for bacon and eggs. No mathematical assumptions were necessary, and the indifference map is not given mathematical form. Some judgments were partly or completely inconsistent with the final map, but not too many. The only conclusion which this experiment justifies is that it is possible to derive such a composite indifference map.

The final attempt to measure an indifference curve is a very recent one by the psychologists Coombs and Milholland (49). The indifference curve involved is one between risk and value of an object, and so will be discussed below in the section on the theory of risky decisions. It is mentioned here because the same methods (which show only that the indifference curve is convex to the origin, and so perhaps should not be called measurement) could equally well be applied to the determination of indifference curves in riskless situations.

Mention should be made of the extensive economic work on statistical demand curves. For some reason the most distinguished statistical demand curve derivers feel it necessary to give an account of consumer's choice theory as a preliminary to the derivation of their empirical demand curves. The result is that the two best books in the area (172, 182) are each divided into two parts; the first is a general discussion of the theory of consumer's choice and the second a quite unrelated report of statistical economic work. Stigler (179) has given good reasons why the statistical demand curves are so little related to the demand curves of economic theory, and Wallis and Friedman (200) argue plausibly that this state

---

of affairs is inevitable. At any rate, there seems to be little prospect of using large-scale economic data to fill in the empirical content of the theory of individual decision making.

Psychological comments. There are several commonplace observations that are likely to occur to psychologists as soon as they try to apply the theory of riskless choices to actual experimental work. The first is that human beings are neither perfectly consistent nor perfectly sensitive. This means that indifference curves are likely to be observable as indifference regions, or as probability distributions of choice around a central locus. It would be easy to assume that each indifference curve represents the modal value of a normal sensitivity curve, and that choices should have statistical properties predictable from that hypothesis as the amounts of the commodities (locations in product space) are changed. This implies that the definition of indifference between two collections of commodities should be that each collection is preferred over the other 50 per cent of the time. Such a definition has been proposed by an economist (108), and used in experimental work by psychologists (142). Of course, 50 per cent choice has been a standard psychological definition of indifference since the days of Fechner.

Incidentally, failure on the part of an economist to understand that a just noticeable difference (j.n.d.) is a statistical concept has led him to argue that the indifference relation is intransitive, that is, that if A is indifferent to B and B is indifferent to C, then A need not be indifferent to C (8, 9, 10). He argues that if A and B are less than one j.n.d. apart, then A will be indifferent to B; the same of course is true of B and C; but A and

C may be more than one j.n.d. apart, and so one may be preferred to the other. This argument is, of course, wrong. If A has slightly more utility than B, then the individual will choose A in preference to B slightly more than 50 per cent of the time, even though A and B are less than one j.n.d. apart in utility. The 50 per cent point is in theory a precisely defined point, not a region. It may in fact be difficult to determine because of inconsistencies in judgments and because of changes in taste with time.

The second psychological observation is that it seems impossible even to dream of getting experimentally an indifference map in n-dimensional space where n is greater than 3. Even the case of $n = 3$ presents formidable experimental problems. This is less important to the psychologist who wants to use the theory of choice to rationalize experimental data than to the economist who wants to derive a theory of general static equilibrium.

Experiments like Thurstone's (185) involve so many assumptions that it is difficult to know what their empirical meaning might be if these assumptions were not made. Presumably, the best thing to do with such experiments is to consider them as tests of the assumption with the least face validity. Thurstone was willing to assume utility maximization and independence of the commodities involved (incidentally, his choice of commodities seems singularly unfortunate for justifying an assumption of independent utilities), and so used his data to construct a utility function. Of course, if only ordinal utility is assumed, then experimental indifference curves cannot be used this way. In fact, in an ordinal-utility universe neither of the principal assumptions made by Thurstone

can be tested by means of experimental indifference curves. So the ... use of cardinal utility, though not necessary, seems to lead to considerably more specific uses for experimental data.

At any rate, from the experimental point of view the most interesting question is: What is the observed shape of indifference curves between independent commodities? This question awaits an experimental answer.

The notion of utility is very similar to the Lewinian notion of valence (120, 121). Lewin conceives of valence as the attractiveness of an object or activity to a person (121). Thus, psychologists might consider the experimental study of utilities to be the experimental study of valences, and therefore an attempt at quantifying parts of the Lewinian theoretical schema.

APPLICATION OF THE THEORY OF RISKLESS CHOICES TO WELFARE ECONOMICS[4]

The classical utility theorists assumed the existence of interpersonally comparable cardinal utility. They were thus able to find a simple answer to the question of how to determine the best economic policy: That economic policy is best which results in the maximum total utility, summed over all members of the economy.

The abandonment of interpersonal comparability makes this answer useless. A sum is meaningless if the units being summed are of varying sizes and there is no way of reducing them to some common size. This

[4] The discussion of welfare economics given in this paper is exceedingly sketchy. For a picture of what the complexities of modern welfare economics are really like (see 11, 13, 14, 86, 118, 124, 127, 139, 140, 148, 154, 155, 166, 174).

point has not been universally recognized, and certain economists (e.g., 82, 154) still defend cardinal (but not interpersonally comparable) utility on grounds of its necessity for welfare economics.

Pareto's principle. The abandonment of interpersonal comparability and then of cardinal utility produced a search for some other principle to justify economic policy. Pareto (146), who first abandoned cardinal utility, provided a partial solution. He suggested that a change should be considered desirable if it left everyone at least as well off as he was before, and made at least one person better off.

Compensation principle. Pareto's principle is fine as far as it goes, but it obviously does not go very far. The economic decisions which can be made on so simple a principle are few and insignificant. So welfare economics languished until Kaldor (98) proposed the compensation principle. This principle is that if it is possible for those who gain from an economic change to compensate the losers for their losses and still have something left over from their gains, then the change is desirable. Of course, if the compensation is actually paid, then this is simply a case of Pareto's principle.

But Kaldor asserted that the compensation need not actually be made; all that was necessary was that it could be made. The fact that it could be made, according to Kaldor, is evidence that the change produces an excess of good over harm, and so is desirable. Scitovsky (173) observed an inconsistency in Kaldor's position: Some cases could arise in which, when a change from A to B has been made because of Kaldor's criterion, then a change back from B to A would also satisfy Kaldor's

criterion. It is customary, therefore, to assume that changes which meet the original Kaldor criterion are only desirable if the reverse change does not also meet the Kaldor criterion.

It has gradually become obvious that the Kaldor-Scitovsky criterion does not solve the problem of welfare economics (see e.g., 18, 99). It assumes that the unpaid compensation does as much good to the person who gains it as it would if it were paid to the people who lost by the change. For instance, suppose that an industrialist can earn $10,000 a year more from his plant by using a new machine, but that the introduction of the machine throws two people irretrievably out of work. If the salary of each worker prior to the change was $4,000 a year, then the industrialist could compensate the workers and still make a profit. But if he does not compensate the workers, then the added satisfaction he gets from his extra $10,000 may be much less than the misery he produces in his two workers. This example only illustrates the principle; it does not make much sense in these days of progressive income taxes, unemployment compensation, high employment, and strong unions.

Social welfare functions. From here on the subject of welfare economics gets too complicated and too remote from psychology to merit extensive exploration in this paper. The line that it has taken is the assumption of a social welfare function (21), a function which combines individual utilities in a way which satisfies Pareto's principle but is otherwise undefined. In spite of its lack of definition, it is possible to draw certain conclusions from such a function (see e.g., 164). However, Arrow (14) has recently shown that a social welfare function that meets certain

very reasonable requirements about being sensitive in some way to the wishes of all the people affected, etc., cannot in general be found in the absence of interpersonally comparable utilities (see also 89).

Psychological comment. Some economists are willing to accept the fact that they are inexorably committed to making moral judgments when they recommend economic policies (e.g., 152, 153). Others still long for the impersonal amorality of a utility measure (e.g., 154). However desirable interpersonally comparable cardinal utility may be, it seems utopian to hope that any experimental procedure will ever give information about individual utilities that could be of any practical use in guiding large-scale economic policy.

THE THEORY OF RISKY CHOICES[5]

Risk and uncertainty. Economists and statisticians distinguish between

[5] Strotz (183) and Alchian (1) present nontechnical and sparkling expositions of the von Neumann and Morgenstern utility measurement proposals. Georgescu-Roegen (78) critically discusses various axiom systems so as to bring some of the assumptions underlying this kind of cardinal utility into clear focus. Allais (3) reviews some of these ideas in the course of criticizing them. Arrow (12, 14) reviews parts of the field.

There is a large psychological literature on one kind of risky decision making, the kind which results when psychologists use partial reinforcement. This literature has been reviewed by Jenkins and Stanley (96). Recently a number of experimenters, including Jarrett (95), Flood (69, 70), Bilodeau (27), and myself (56) have been performing experiments on human subjects who are required to choose repetitively between two or more alternatives, each of which has a probability of reward greater than zero and less than one. The problems raised by these experiments are too complicated and too far removed from conventional utility theory to be dealt with in this paper. This line of experimentation may eventually provide the link which ties together utility theory and reinforcement theory.

risk and uncertainty. There does not seem to be any general agreement about which concept should be associated with which word, but the following definitions make the most important distinctions.

Almost everyone would agree that when I toss a coin the probability that I will get a head is .5. A proposition about the future to which a number can be attached, a number that represents the likelihood that the proposition is true, may be called a *first-order risk*. What the rules are for attaching such numbers is a much debated question, which will be avoided in this paper.

Some propositions may depend on more than one probability distribution. For instance, I may decide that if I get a tail, I will put the coin back in my pocket, whereas if I get a head, I will toss it again. Now, the probability of the proposition "I will get a head on my second toss" is a function of two probability distributions, the distribution corresponding to the first toss and that corresponding to the second toss. This might be called a *second-order risk*. Similarly, risks of any order may be constructed. It is a mathematical characteristic of all higher-order risks that they may be compounded into first-order risks by means of the usual theorems for compounding probabilities. (Some economists have argued against this procedure [83], essentially on the grounds that you may have more information by the time the second risk comes around. Such problems can best be dealt with by means of von Neumann and Morgenstern's [197] concept of strategy, which is discussed below. They become in general problems of uncertainty, rather than risk.)

Some propositions about the future exist to which no generally accepted probabilities can be attached. What is the probability that the following proposition is true: Immediately after finishing this paper, you will drink a glass of beer? Surely it is neither impossible nor certain, so it ought to have a probability between zero and one, but it is impossible for you or me to find out what that probability might be, or even to set up generally acceptable rules about how to find out. Such propositions are considered cases of *uncertainty*, rather than of risk. This section deals only with the subject of first-order risks. The subject of uncertainty will arise again in connection with the theory of games.

Expected utility maximization. The traditional mathematical notion for dealing with games of chance (and so with risky decisions) is the notion that choices should be made so as to maximize *expected value*. The expected value of a bet is found by multiplying the value of each possible outcome by its probability of occurrence and summing these products across all possible outcomes. In symbols:

$$EV = p_1\$_1 + p_2\$_2 + \cdots + p_n\$_n,$$

where p stands for probability, $\$$ stands for the value of an outcome, and $p_1 + p_2 + \cdots + p_n = 1$.

The assumption that people actually behave the way this mathematical notion says they should is contradicted by observable behavior in many risky situations. People are willing to buy insurance, even though the person who sells the insurance makes a profit. People are willing to buy lottery tickets, even though the lottery makes a profit. Consideration of the problem of insurance and of the St. Petersburg paradox led Daniel Bernoulli, an eighteenth century mathematician, to propose that they could be resolved by assuming that

people act so as to maximize *expected utility*, rather than expected value (26). (He also assumed that utility followed a function that more than a century later was proposed by Fechner for subjective magnitudes in general and is now called Fechner's Law.) This was the first use of the notion of expected utility.

The literature on risky decision making prior to 1944 consists primarily of the St. Petersburg paradox and other gambling and probability literature in mathematics, some literary discussion in economics (e.g., 109, 187), one economic paper on lotteries (189), and the early literature of the theory of games (31, 32, 33, 34, 195), which did not use the notion of utility. The modern period in the study of risky decision making began with the publication in 1944 of von Neumann and Morgenstern's monumental book *Theory of Games and Economic Behavior* (196, see also 197), which we will discuss more fully later. Von Neumann and Morgenstern pointed out that the usual assumption that economic man can always say whether he prefers one state to another or is indifferent between them needs only to be slightly modified in order to imply cardinal utility. The modification consists of adding that economic man can also completely order probability combinations of states. Thus, suppose that an economic man is indifferent between the certainty of $7.00 and a 50-50 chance of gaining $10.00 or nothing. We can assume that his indifference between these two prospects means that they have the same utility for him. We may define the utility of $0.00 as zero utiles (the usual name for the unit of utility, just as sone is the name for the unit of auditory loudness), and the utility of $10.00 as 10 utiles. These two

arbitrary definitions correspond defining the two undefined constants which are permissible since cardinal utility is measured only up to a linear transformation. Then we may calculate the utility of $7.00 by using the concept of expected utility as follows:

$$U(\$7.00) = .5\,U(\$10.00) + .5\,U(\$0.00)$$
$$= .5(10) + .5(0) = 5.$$

Thus we have determined the cardinal utility of $7.00 and found that it is 5 utiles. By varying the probabilities and by using the already found utilities it is possible to discover the utility of any other amount of money, using only the two permissible arbitrary definitions. It is even more convenient if instead of $+$10.00, $-$10.00 or some other loss is used as one of the arbitrary utilities.

A variety of implications is embodied in this apparently simple notion. In the attempt to examine and exhibit clearly what these implications are, a number of axiom systems, differing from von Neumann and Morgenstern's but leading to the same result, have been developed (73, 74, 85, 135, 136, 171). This paper will not attempt to go into the complex discussions (e.g., 130, 131, 168, 207) of these various alternative axiom systems. One recent discussion of them (78) has concluded, on reasonable grounds, that the original von Neumann and Morgenstern set of axioms is still the best.

It is profitable, however, to examine what the meaning of this notion is from the empirical point of view if it is right. First, it means that risky propositions can be ordered in desirability, just as riskless ones can. Second, it means that the concept of expected utility is behaviorally meaningful. Finally, it means that choices among risky alternatives

are made in such a way that they maximize expected utility.

If this model is to be used to predict actual choices, what could go wrong with it? It might be that the probabilities by which the utilities are multiplied should not be the objective probabilities; in other words, a decider's estimate of the subjective importance of a probability may not be the same as the numerical value of that probability. It might be that the method of combination of probabilities and values should not be simple multiplication. It might be that the method of combination of the probability-value products should not be simple addition. It might be that the process of gambling has some positive or negative utility of its own. It might be that the whole approach is wrong, that people just do not behave as if they were trying to maximize expected utility. We shall examine some of these possibilities in greater detail below.

Economic implications of maximizing expected utility. The utility-measurement notions of von Neumann and Morgenstern were enthusiastically welcomed by many economists (e.g., **73, 193**), though a few (e.g., **19**) were at least temporarily (**20**) unconvinced. The most interesting economic use of them was proposed by Friedman and Savage (**73**), who were concerned with the question of why the same person who buys insurance (with a negative expected money value), and therefore is willing to pay in order not to take risks, will also buy lottery tickets (also with a negative expected money value) in which he pays in order to take risks. They suggested that these facts could be reconciled by a doubly inflected utility curve for money, like that in Fig. 2. If *I* represents the person's current income, then he is clearly willing to accept "fair" insurance (i.e., insurance with zero expected money value) because the serious loss against which he is insuring would have a lower expected utility than the certain loss of the insurance premium. (Negatively accelerated total utility curves, like that from the origin to *I*, are what you get when marginal utility decreases; thus, decreasing marginal
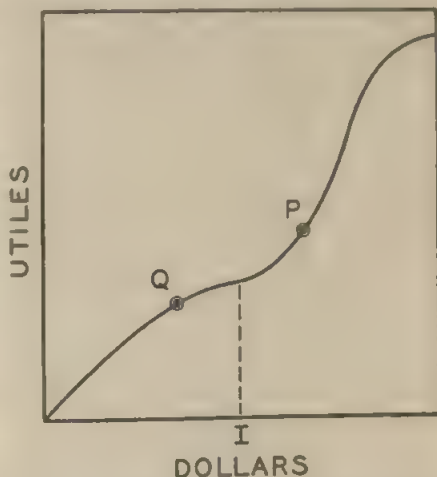
FIG. 2. HYPOTHETICAL UTILITY CURVE FOR MONEY, PROPOSED BY FRIEDMAN AND SAVAGE

utility is consistent with the avoidance of risks.) The person would also be willing to buy lottery tickets, since the expected utility of the lottery ticket is greater than the certain loss of the cost of the ticket, because of the rapid increase in the height of the utility function. Other considerations make it necessary that the utility curve turn down again. Note that this discussion assumes that gambling has no inherent utility.

Markowitz (**132**) suggested an important modification in this hypothesis. He suggested that the origin of a person's utility curve for money be taken as his customary

financial status, and that on both sides of the origin the curve be assumed first concave and then convex. If the person's customary state of wealth changes, then the shape of his utility curve will thus remain generally the same with respect to where he now is, and so his risk-taking behavior will remain pretty much the same instead of changing with every change of wealth as in the Friedman-Savage formulation.

Criticism of the expected-utility maximization theory. It is fairly easy to construct examples of behavior that violate the von Neumann-Morgenstern axioms (for a particularly ingenious example, see 183). It is especially easy to do so when the amounts of money involved are very large, or when the probabilities or probability differences involved are extremely small. Allais (5) has constructed a questionnaire full of items of this type. For an economist interested in using these axioms as a basis for a completely general theory of risky choice, these examples may be significant. But psychological interest in this model is more modest. The psychologically important question is: Can such a model be used to account for simple experimental examples of risky decisions?

Of course a utility function derived by von Neumann-Morgenstern means is not necessarily the same as a classical utility function (74, 203; see also 82).

Experiment on the von Neumann-Morgenstern model. A number of experiments on risky decision making have been performed. Only the first of them, by Mosteller and Nogee (142), has been in the simple framework of the model described above. All the rest have in some way or another centered on the concept of probabilities effective for behavior

which differ in some way from the objective probabilities, as well as on utilities different from the objective values of the objects involved.

Mosteller and Nogee (142) carried out the first experiment to apply the von Neumann-Morgenstern model. They presented Harvard undergraduates and National Guardsmen with bets stated in terms of rolls at poker dice, which each subject could accept or refuse. Each bet gave a "hand" at poker dice. If the subject could beat the hand, he won an amount stated in the bet. If not, he lost a nickel. Subjects played with $1.00, which they were given at the beginning of each experimental session. They were run together in groups of five; but each decided and rolled the poker dice for himself. Subjects were provided with a table in which the mathematically fair bets were shown, so that a subject could immediately tell by referring to the table whether a given bet was fair, or better or worse than fair.

In the data analysis, the first step was the determination of "indifference offers." For each probability used and for each player, the amount of money was found for which that player would accept the bet 50 per cent of the time. Thus equality was defined as 50 per cent choice, as it is likely to be in all psychological experiments of this sort. Then the utility of $0.00 was defined as 0 utiles, and the utility of losing a nickel was defined as -1 utile. With these definitions and the probabilities involved, it was easy to calculate the utility corresponding to the amount of money involved in the indifference offer. It turned out that, in general, the Harvard undergraduates had diminishing marginal utilities, while the National Guardsmen had increasing marginal utilities.

The utilities thus calculated were used in predicting the results of more complex bets. It is hard to evaluate the success of these predictions. At any rate, an auxiliary paired-comparisons experiment showed that the hypothesis that subjects maximized expected utility predicted choices better than the hypothesis that subjects maximized expected money value.

The utility curve that Mosteller and Nogee derive is different from the one Friedman and Savage (73) were talking about. Suppose that a subject's utility curve were of the Friedman-Savage type, as in Fig. 2, and that he had enough money to put him at point *P*. If he now wins or loses a bet, then he is moved to a different location on the indifference curve, say *Q*. (Note that the amounts of money involved are much smaller than in the original Friedman-Savage use of this curve.) However, the construction of a Mosteller-Nogee utility curve assumes that the individual is always at the same point on his utility curve, namely the origin. This means that the curve is really of the Markowitz (132) type discussed above, instead of the Friedman-Savage type. The curve is not really a curve of utility of money in general, but rather it is a curve of the utility-for-*n*-more dollars. Even so, it must be assumed further that as the total amount of money possessed by the subject changes during the experiment, the utility-for-*n*-more dollars curve does not change. Mosteller and Nogee argue, on the basis of detailed examination of some of their data, that the amount of money possessed by the subjects did not seriously influence their choices. The utility curves they reported showed changing marginal utility within the amounts of money used in their ex-

periment. Consequently, their conclusion that the amount of money possessed by the subjects was not seriously important can only be true if their utility curves are utility-for-*n*-more dollars curves and if the shapes of such curves are not affected by changes in the number of dollars on hand. This discussion exhibits a type of problem which must always arise in utility measurement and which is new in psychological scaling. The effects of previous judgments on present judgments are a familiar story in psychophysics, but they are usually assumed to be contaminating influences that can be minimized or eliminated by proper experimental design. In utility scaling, the fundamental idea of a utility scale is such that the whole structure of a subject's choices should be altered as a result of each previous choice (if the choices are real ones involving money gains or losses). The Markowitz solution to this problem is the most practical one available at present, and that solution is not entirely satisfactory since all it does is to assume that people's utilities for money operate in such a way that the problem does not really exist. This assumption is plausible for money, but it gets rapidly less plausible when other commodities with a less continuous character are considered instead.

Probability preferences. In a series of recent experiments (55, 57, 58, 59), the writer has shown that subjects, when they bet, prefer some probabilities to others (57), and that these preferences cannot be accounted for by utility considerations (59). All the experiments were basically of the same design. Subjects were required to choose between pairs of bets according to the method of paired comparisons. The bets were of three kinds: positive expected value, nega-

tive expected value, and zero expected value. The two members of each pair of bets had the same expected value, so that there was never (in the main experiment [57, 59]) any objective reason to expect that choosing one bet would be more desirable than choosing the other.

Subjects made their choices under three conditions: just imagining they were betting; betting for worthless chips; and betting for real money. They paid any losses from their own funds, but they were run in extra sessions after the main experiment to bring their winnings up to $1.00 per hour.

The results showed that two factors were most important in determining choices: general preferences or dislikes for risk-taking, and specific preferences among probabilities. An example of the first kind of factor is that subjects strongly preferred low probabilities of losing large amounts of money to high probabilities of losing small amounts of money—they just didn't like to lose. It also turned out that on positive expected value bets, they were more willing to accept long shots when playing for real money than when just imagining or playing for worthless chips. An example of the second kind of factor is that they consistently preferred bets involving a 4/8 probability of winning to all others, and consistently avoided bets involving a 6/8 probability of winning. These preferences were reversed for negative expected value bets.

These results were independent of the amounts of money involved in the bets, so long as the condition of constant expected value was maintained (59). When pairs of bets which differed from one another in expected value were used, the choices were a compromise between maximizing expected amount of money and betting at the preferred probabilities (58). An attempt was made to construct individual utility curves adequate to account for the results of several subjects. For this purpose, the utility of $0.30 was defined as 30 utiles, and it was assumed that subjects cannot discriminate utility differences smaller than half a utile. Under these assumptions, no individual utility curves consistent with the data could be drawn. Various minor experiments showed that these results were reliable and not due to various possible artifacts (59). No attempt was made to generate a mathematical model of probability preferences.

The existence of probability preferences means that the simple von Neumann-Morgenstern method of utility measurement cannot succeed. Choices between bets will be determined not only by the amounts of money involved, but also by the preferences the subjects have among the probabilities involved. Only an experimental procedure which holds one of these variables constant, or otherwise allows for it, can hope to measure the other. Thus my experiments cannot be regarded as a way of measuring probability preferences; they show only that such preferences exist.

It may nevertheless be possible to get an interval scale of the utility of money from gambling experiments by designing an experiment which measures utility and probability preferences simultaneously. Such experiments are likely to be complicated and difficult to run, but they can be designed.

Subjective probability. First, a clarification of terms is necessary. The phrase *subjective probability* has been used in two ways: as a name for a school of thought about the

logical basis of mathematical probability (51, 52, 80) and as a name for a transformation on the scale of mathematical probabilities which is somehow related to behavior. Only the latter usage is intended here. The clearest distinction between these two notions arises from consideration of what happens when an objective probability can be defined (e.g., in a game of craps). If the subjective probability is assumed to be different from the objective probability, then the concept is being used in its second, or psychological, sense. Other terms with the same meaning have also been used: personal probability, psychological probability, expectation (a poor term because of the danger of confusion with expected value). (For a more elaborate treatment of concepts in this area, see 192.)

In 1948, prior to the Mosteller and Nogee experiment, Preston and Baratta (149) used essentially similar logic and a somewhat similar experiment to measure subjective probabilities instead of subjective values. They required subjects to bid competitively for the privilege of taking a bet. All bids were in play money, and the data consisted of the winning bids. If each winning bid can be considered to represent a value of play money such that the winning bidder is indifferent between it and the bet he is bidding for, and if it is further assumed that utilities are identical with the money value of the play money and that all players have the same subjective probabilities, then these data can be used to construct a subjective probability scale. Preston and Baratta constructed such a scale. The subjects, according to the scale, overestimate low probabilities and underestimate high ones, with an indifference point (where subjective probability equals objective probability) at about 0.2. Griffith (81) found somewhat similar results in an analysis of parimutuel betting at race tracks, as did Attneave (17) in a guessing game, and Sprowls (178) in an analysis of various lotteries. The Mosteller and Nogee data (142) can, of course, be analyzed for subjective probabilities instead of subjective values. Mosteller and Nogee performed such an analysis and said that their results were in general agreement with Preston and Baratta's. However, Mosteller and Nogee found no indifference point for their Harvard students, whereas the National Guardsmen had an indifference point at about 0.5. They are not able to reconcile these differences in results.

The notion of subjective probability has some serious logical difficulties. The scale of objective probability is bounded by 0 and 1. Should a subjective probability scale be similarly bounded, or not? If not, then many different subjective probabilities will correspond to the objective probabilities 0 and 1 (unless some transformation is used so that 0 and 1 objective probabilities correspond to infinite subjective probabilities, which seems unlikely). Considerations of the addition theorem to be discussed in a moment have occasionally led people to think of a subjective probability scale bounded at 0 but not at 1. This is surely arbitrary. The concept of absolute certainty is neither more nor less indeterminate than is the concept of absolute impossibility.

Even more drastic logical problems arise in connection with the addition theorem. If the objective probability of event A is P, and that of A not occurring is Q, then $P + Q = 1$. Should this rule hold for subjective probabilities? Intuitively it seems neces-

sary that if we know the subjective probability of A, we ought to be able to figure out the subjective probability of not-A, and the only reasonable rule for figuring it out is subtraction of the subjective probability of A from that of complete certainty. But the acceptance of this addition theorem for subjective probabilities plus the idea of bounded subjective probabilities means that the subjective probability scale must be identical with the objective probability scale. Only for a subjective probability scale identical with the objective probability scale will the subjective probabilities of a collection of events, one of which must happen, add up to 1. In the special case where only two events, A and not-A, are considered, a subjective probability scale like $S1$ or $S2$ in Fig. 3 would meet the requirements of additivity, and this fact has led to some speculation about such scales, particularly about $S1$. But such scales do not meet the additivity requirements when more than two events are considered.

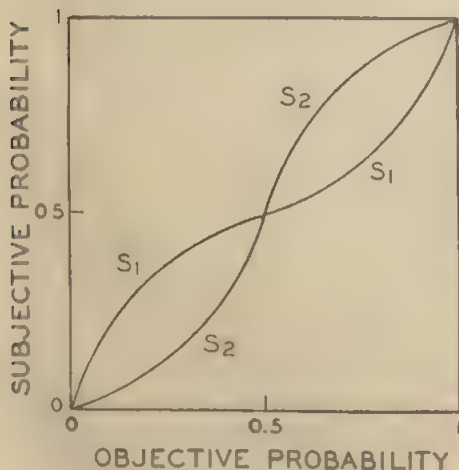One way of avoiding these diffi-

FIG. 3. HYPOTHETICAL SUBJECTIVE PROBABILITY CURVES

culties is to stop thinking about a scale of subjective probabilities and, instead, to think of a weighting function applied to the scale of objective probabilities which weights these objective probabilities according to their ability to control behavior. Presumably, I was studying this ability in my experiments on probability preferences (55, 57, 58, 59). There is no reason why such weighted probabilities should add up to 1 or should obey any other simple combinatory principle.

Views and experiments which combine utility and subjective probability. The philosopher Ramsey published in 1926 (reprinted in 150) an essay on the subjective foundations of the theory of probability; this contained an axiom system in which both utility and subjective probability appeared. He used 0.5 subjective probability as a reference point from which to determine utilities, and then used these utilities to determine other subjective probabilities. Apparently, economists did not discover Ramsey's essay until after von Neumann and Morgenstern's book aroused interest in the subject. The only other formal axiom system in which both utility and subjective probability play a part is one proposed by Savage (171), which is concerned with uncertainty, rather than risk, and uses the concept of subjective probability in its theory-of-probability sense.

The most extensive and important experimental work in the whole field of decision making under risk and uncertainty is now being carried out by Coombs and his associates at the University of Michigan. Coombs's thinking about utility and subjective probability is an outgrowth of his thinking about psychological scaling in general. (For a discussion of his views, see 43, 44, 45, 46, 47.) The

essence of his work is the attempt to measure both utility and subjective probability on an ordered metric scale. An ordered metric scale has all the properties of an ordinal scale, and, in addition, the distances between some or all of the stimuli can be rank ordered. Coombs has developed various experimental procedures for obtaining such information about the spacings of stimuli.

In the most important article on utility and subjective probability to come out of the Coombs approach, Coombs and Beardslee (48) present an analysis of gambling decisions involving three independent variables: utility for prize, utility for stake, and subjective probability. All three are assumed measurable only up to an ordered metric, although it is assumed that the psychological probability of losing the stake is one minus the psychological probability of winning the prize, an assumption that limits the permissible underlying psychological probability functions to shapes like those in Fig. 3. An elaborate graphic analysis of the indifference surfaces in this three-dimensional space is given, containing far too many interesting relationships to summarize here. An experiment based on this model was designed. Coombs is reluctant to use sums of money as the valuable objects in his experiments because of the danger that subjects will respond to the numerical value of the amount of dollars rather than to the psychological value. Therefore he used various desirable objects (e.g., a radio) as stimuli, and measured their utility by the techniques he has developed to obtain ordered metric scales. He used simple numerical statements of probability as the probability stimuli, and assumed that subjective probability was equal to

objective probability. The subject from whose judgments the ordered metric utility measurement was constructed was then presented with imaginary bets involving these objects and probabilities, and it turned out that she almost always chose the one with the higher expected utility. This experiment is significant only as an illustration of the application of the method; the conclusion that subjects attempt to maximize expected utility cannot very comfortably be generalized to other subjects and to real choices without better evidence.

Coombs and Milholland (49) did a much more elaborate experiment in which they established ordered metric scales, both for the utilities of a collection of objects and for the subjective probabilities of a collection of statements (e.g., Robin Roberts will win 20 games next year). Statements and objects were combined into "bets," and the two subjects for whom the ordered metric scales had been established were asked to make judgments about which bet they would most, and which they would least, prefer from among various triads of bets. These judgments were examined to discover whether or not they demonstrated the existence of at least one convex indifference curve between utility and subjective probability (the requirements for demonstrating the convexity of an indifference curve by means of ordered metric judgments are fairly easy to state). A number of cases consistent with a convex indifference curve were found, but a retest of the ordered metric data revealed changes which eliminated all of the cases consistent with a convex indifference curve for one subject, and all but one case for the other. It is not possible to make a statistical test of whether or not

that one case might have come about by chance. No evidence was found for the existence of concave indifference curves, which are certainly inconsistent with the theory of risky decisions. This experiment is a fine example of the strength and weakness of the Coombs approach. It makes almost no assumptions, takes very little for granted, and avoids the concept of error of judgment; as a result, much of the potential information in the data is unused and rarely can any strong conclusions be drawn.

A most disturbing possibility is raised by experiments by Marks (133) and Irwin (94) which suggest that the shape of the subjective probability function is influenced by the utilities involved in the bets. If utilities and subjective probabilities are not independent, then there is no hope of predicting risky decisions unless their law of combination is known, and it seems very difficult to design an experiment to discover that law of combination. However, the main differences that Marks and Irwin found were between probabilities attached to desirable and undesirable alternatives. It is perfectly possible that there is one subjective probability function for bets with positive expected values and a different one for bets with negative expected values, just as the negative branch of the Markowitz utility function is likely to be different from the positive branch. The results of my probability preference experiments showed very great differences between the probability preference patterns for positive and for negative expected-value bets (57), but little difference between probability preferences at different expected-value levels so long as zero expected value was not crossed (59). This evidence supports the idea that perhaps only two subjective probability functions are necessary.

Santa Monica Seminar. In the summer of 1952 at Santa Monica, California, a group of scientists conferred on problems of decision making. They met in a two-month seminar sponsored by the University of Michigan and the Office of Naval Research. The dittoed reports of these meetings are a gold mine of ideas for the student of this problem. Some of the work done at this seminar is now being prepared for a book on *Decision Processes* edited by R. M. Thrall, C. H. Coombs, and R. L. Davis, of the University of Michigan.

Several minor exploratory experiments were done at this seminar. Vail (190) did an experiment in which he gave four children the choice of which side of various bets they wanted to be on. On the assumption of linear utilities, he was able to compute subjective probabilities for these children. The same children, however, were used as subjects for a number of other experiments; so, when Vail later tried them out on some other bets, he found that they consistently chose the bet with the highest probability of winning, regardless of the amounts of money involved. When 50-50 bets were involved, one subject consistently chose the bet with the *lowest* expected value. No generalizable conclusions can be drawn from these experiments.

Kaplan and Radner (100) tried out a questionnaire somewhat like Coombs's method of measuring subjective probability. Subjects were asked to assign numbers to various statements. The numbers could be anything from 0 to 100 and were to represent the likelihood that the statement was true. The hypotheses to be tested were: (a) for sets of ex-

taustive and mutually exclusive
statements in which the numbers as-
signed (estimates of degree of belief)
were nearly equal, the sums of these
numbers over a set would increase
with the number of alternatives (be-
cause low probabilities would be over-
estimated); (*b*) for sets with the same
numbers of alternatives, those with
one high number assigned would have
a lower set sum than those with no
high numbers. The first prediction
was verified; the second was not.
Any judgments of this sort are so
much more likely to be made on the
basis of number preferences and
similar variables than on subjective
probabilities that they offer very
little hope as a method of measuring
subjective probabilities.

Variance preferences. Allais (**2, 3,
4**) and Georgescu-Roegen (**78**) have
argued that it is not enough to apply
a transform on objective value and on
objective probability in order to pre-
dict risky decisions from expected
utility (see also **188**); it is also neces-
sary to take into account at least the
variance, and possibly the higher
moments, of the utility distribution.
There are instances in which this
argument seems convincing. You
would probably prefer the certainty
of a million dollars to a 50-50 chance
of getting either four million or noth-
ing. I do not think that this prefer-
ence is due to the fact that the ex-
pected utility of the 50-50 bet is less
than the utility of one million dollars
to you, although this is possible. A
more likely explanation is simply
that the variances of the two propo-
sitions are different. Evidence in
favor of this is the fact that if you
knew you would be offered this choice
20 times in succession, you would
probably take the 50-50 bet each
time. Allais (**5**) has constructed a
number of more sophisticated exam-

ples of this type. However, from a
simple-minded psychological point of
view, these examples are irrelevant.
It is enough if the theory of choice
can predict choices involving familiar
amounts of money and familiar
probability differences—choices such
as those which people are accustomed
to making. It may be necessary for
economic theory that the theory of
choice be universal and exceptionless,
but experimental psychologists need
not be so ambitious. This is fortu-
nate, because the introduction of the
variance and higher moments of the
utility distribution makes the prob-
lem of applying the theory experi-
mentally seem totally insoluble. It is
difficult enough to derive reasonable
methods of measuring utility alone
from risky choices; when it also be-
comes necessary to measure subjec-
tive probability and to take the
higher moments of the utility dis-
tribution into account, the problem
seems hopeless. Allais apparently
hopes to defeat this problem by using
psychophysical methods to measure
utility (and presumably subjective
probability also). This is essentially
what Coombs has done, but Coombs
has recognized that such procedures
are unlikely to yield satisfactory
interval scales. The dollar scale of
the value of money is so thoroughly
taught to us that it seems almost im-
possible to devise a psychophysical
situation in which subjects would
judge the utility, rather than the dol-
lar value, of dollars. They might
judge the utility of other valuable
objects, but since dollars are the
usual measure of value, such judg-
ments would be less useful, and even
these judgments would be likely to be
contaminated by the dollar values of
the objects. I would get more utility
from a new electric shaver than I
would from a new washing machine,

but because of my knowledge of the relative money values of these objects, I would certainly choose the washing machine if given a choice between them. Somewhat similar arguments can be applied against using psychophysical methods to measure subjective probability. A final point is that, since these subjective scales are to be used to predict choices, it would be best if they could be derived from similar choices.

Other approaches. Shackle (175) has proposed a theory of decision making under risk and uncertainty. This theory is unique in that it does not assume any kind of maximizing behavior. For every possible outcome of a decision made in a risky or uncertain situation, Shackle assumes that there is a degree of potential surprise that this, rather than some other, outcome would occur. Every outcome-potential surprise pair is ranked in accordance with its ability to stimulate the mind (stimulation increases with increasing outcome and decreases with increasing potential surprise). The highest-ranking positive outcome-potential surprise pair and the highest-ranking negative pair are found, and these two possibilities alone determine what the individual will do. Semi-mathematical methods are used to predict the outcome of consideration of possible lines of action. Although attempts have been made to relate it to Wald's minimax principle for statistical decision functions (see below), the fact remains that most critics of the Shackle point of view have judged it to be either too vague to be useful, or, if specified in detail, too conducive to patently absurd predictions (e.g., 201).

Shackle's point of view was developed primarily to deal with unique choices—choices which can be made only once. Allais (3) has similarly

criticized conventional utility theory's attack on this problem. Since the usual frequency theory of probability conceives of the probability as the limit of the outcomes of a large number of similar trials, it is questionable that notions which use probability in the ordinary sense (like the notion of maximizing expected utility) are applicable to unique choices. However, this seems to be an experimental problem. If notions which use ordinary probability are incapable of predicting actual unique choices, then it will be necessary to seek other theoretical tools. But so long as a generally acceptable probability can be defined (e.g., as in the unique toss of a coin), it is not necessary to assume a priori that theories based on conventional probabilities will be inadequate. When no generally acceptable probability can be defined, then the problem becomes very different.

Cartwright and Festinger (38, 41) have proposed a theory about the time it takes to make decisions which is in some ways similar to those discussed in this section. The main difference is that they add the concept of restraining forces, and that they conceive of all subjective magnitudes as fluctuating randomly around a mean value. From this they deduce various propositions about decision times and the degree of certainty which subjects will feel about their decisions, and apparently these propositions work out experimentally pretty well (38, 39, 61, 62). The Lewinian theoretical orientation seems to lead to this kind of model: Lewin, Dembo, Festinger, and Sears (122) present a formally similar theory about level of aspiration. Of course, the notion of utility is very similar to the Lewinian notion of valence.

Landahl (115) has presented a mathematical model for risk-taking behavior based on the conceptual neurology of the mathematical biophysics school.

Psychological comments. The area of risky decision making is full of fascinating experimental problems. Of these, the development of a satisfactory scale of utility of money and of subjective probability must come first, since the theory of risky decision making is based on these notions. The criterion for satisfactoriness of these scales must be that they successfully predict choices other than those from which they were derived. To be really satisfactory, it is desirable that they should predict choices in a wide variety of differing situations. Unlike the subjective scales usually found in psychophysics, it is likely that these scales will differ widely from person to person, so a new determination of each scale must be made for each new subject. It can only be hoped that the scales do not change in time to any serious degree; if they do, then they are useless.

Once scales of utility and subjective probability are available, then many interesting questions arise. What about the addition theorem for subjective probabilities? Does gambling itself have utility, and how much? To what extent can these subjective scales be changed by learning? To what degree do people differ, and can these differences be correlated with environmental, historical, or personality differences? Finally, psychologists might be able to shed light on the complex economic problem of interacting utilities of different goods.

The area of risky decision making, like the area of the theory of games, tends to encourage in those interested in it the custom of carrying out small pilot experiments on their sons, laboratory assistants, or secretaries. Such experiments are too seldom adequately controlled, and are almost never used as a basis for larger-scale, well-designed experiments. Whether an ill-designed and haphazardly executed little experiment is better than no experiment at all is questionable. The results of such pilot experiments too often are picked up and written into the literature without adequate warning about the conditions under which they were performed and the consequent limitations on the significance of the results.

THE TRANSITIVITY OF CHOICES

In the section on riskless choices this paper presented a definition of economic man. The most important part of this definition can be summed up by saying that economic man is rational. The concept of rationality involves two parts: that of a weak ordering of preferences, and that of choosing so as to maximize something. Of these concepts, the one which seems most dubious is the one of a weakly ordered preference field. This is dubious because it implies that choices are transitive; that is, if A is preferred to B, and B is preferred to C, then A is preferred to C.

Two economists have designed experiments specifically intended to test the transitivity of choices. Papandreou performed an elaborate and splendidly controlled experiment (145) designed to discover whether or not intransitivities occurred in imagined-choice situations. He prepared triplets of hypothetical bundles of admissions to plays, athletic contests, concerts, etc., and required his subjects to choose between pairs of bundles. Each bundle consisted of a total of four admissions to two events, e.g., 3 plays and 1 tennis

tournament. In the main experiment, each bundle is compared with two others involving the same kinds of events, but in the better designed auxiliary experiment, a total of six different events are used, so that each bundle has no events in common with the other two bundles in its triplet. Since there are three bundles in each triplet, there are three choices between pairs for each triplet, and these choices may, or may not, be transitive. The subjects were permitted to say that they were indifferent between two bundles; consequently there were 27 possible configurations of choices, of which only 13 satisfied the transitivity axiom. In the main experiment, 5 per cent of the triplets of judgments were intransitive; in the auxiliary experiment, only 4 per cent. Papandreou develops a stochastic model for choices under such conditions; the results are certainly consistent with the amount of intransitivity permitted by his model. Papandreou concludes that at least for his specific experimental conditions, transitivity does exist.

May (138), using different kinds of stimuli in a less elaborate experiment, comes up with results less consistent with transitivity. May required a classroom group to make pairwise choices between three marriage partners who were identified only by saying how intelligent, good looking, and rich they were. Judgments of indifference were not permitted. The results were that 27 per cent of the subjects gave intransitive triads of choices. May suggests, very plausibly, that intransitive choices may be expected to occur whenever more than one dimension exists in the stimuli along which subjects may order their preferences. However, May would probably have gotten

fewer intransitivities if he had permitted the indifference judgment. If subjects are really indifferent among all three of the elements of a triad of objects, but are required to choose between them in pairs and do so by chance, then they will choose intransitively one-fourth of the time. Papandreou's stochastic model gives one theory about what happens when preferences diverge just slightly from indifference, but presumably a more detailed model can be worked out. Papandreou's model permits only three states: prefer A to B, prefer B to A, and indifferent. It ought to be possible to base a model for such situations on the cumulative normal curve, and thus to permit any degree of preference. For every combination of degrees of preference, such a model would predict the frequency of intransitive choices.

In the paired comparisons among bets (57) described in the section on risky choices, quite elaborate intransitivities could and did occur. However, it is easy to show that any intransitivity involving four or more objects in a paired comparisons judgment situation will necessarily produce at least one intransitivity involving three objects. Consequently, the intransitive triplet or circular triad is the best unit of analysis for intransitivities in these more complicated judgment situations. I counted the frequency of occurrence of circular triads and found that they regularly occurred about 20 per cent of the total number of times they could occur. (Of course, no indifference judgments could be permitted.) The experiment fulfills May's criterion for the occurrence of intransitivities, since both probability and amount of money were present in each bet, and subjects could be expected to take both into account

when making choices. It might be supposed that the difference between the imaginary choices of the Papandreou and May experiments and the real choices in my experiment would lead to differences in the frequency of occurrence of intransitivities, but there were no substantial differences in my experiment between the frequencies of occurrence in the just-imagining sessions and in the real gambling sessions, and what differences there were, were in the direction of greater transitivity when really gambling. These facts should facilitate further experiments on this problem.

In one sense, transitivity can never be violated. A minimum of three choices is required to demonstrate intransitivity. Since these choices will necessarily be made in sequence, it can always be argued that the person may have changed his tastes between the first choice and the third. However, unless the assumption of constancy of tastes over the period of experimentation is made, no experiments on choice can ever be meaningful, and the whole theory of choice becomes empty (see **184** for a similar situation). So this quibble can be rejected at once.

Utility maximization will not work except with a transitive preference field. Consequently, if the models discussed in this paper are to predict experimental data, it is necessary that intransitivities in these data be infrequent enough to be considered as errors. However, from a slightly different point of view (**54**) the occurrence or nonoccurrence of transitive choice patterns is an experimental phenomenon, and presumably a lawful one. May has suggested what that law is: Intransitivities occur when there are conflicting stimulus dimensions along which to judge.

This notion could certainly be tested and made more specific by appropriate experiments.

A final contribution in a related, but different, area is Vail's stochastic utility model (**191**). Vail assumes that choices are dependent on utilities that oscillate in a random manner around a mean value. From this assumption plus a few other reasonable ones, he deduces that if the over-all preference is $1 > 2 > 3$, and if 1 is preferred to 2 more than 2 is preferred to 3, then the frequencies of occurrence of the six possible transitive orderings should be ordered as follows: $123 > 132 > 213 > 312 > 231 > 321$. This result is certainly easy to test experimentally, and sounds plausible.

The Theory of Games and of Decision Functions[6]

This section will not go into the theory of games or into the intimately related subject of statistical decision functions at all thoroughly. These are mathematical subjects of a highly

[6] Marschak (**134**), Hurwicz (**92**), Neisser (**143**), Stone (**181**), and Kaysen (**107**) published reviews of *The Theory of Games and Economic Behavior* which present the fundamental ideas in much simpler language than the original source. Marschak works out in detail the possible solutions of a complicated three-person bargaining game, and thereby illustrates the general nature of a solution. The two volumes of *Contributions to the Theory of Games* (**112**, **113**), plus McKinsey's book on the subject (**129**), provide an excellent bibliography of the mathematical literature. McKinsey's book is an exposition of the fundamental concepts, intended as a textbook, which is simpler than von Neumann and Morgenstern and pursues certain topics further. Wald's book (**198**) is, of course, the classical work on statistical decision functions. Bross's book (**35**) presents the fundamental ideas about statistical decision functions more simply, and with a somewhat different emphasis. Girshick and Blackwell's book (**79**) is expected to be a very useful presentation of the field.

technical sort, with few statements which lend themselves to experimental test. Rather, the purpose of this section is to show how these subjects relate to what has gone before, to give a brief summary of the contents of *Theory of Games and Economic Behavior* by von Neumann and Morgenstern (197), and to describe a few experiments in the area of game playing —experiments which are stimulated by the theory of games although not directly relevant to it.

The theory of games. The theory of games probably originated in the work of Borel (**31, 32, 33, 34**; see also **71, 72**) in the 1920's. In 1928, von Neumann (**195**), working independently of Borel, published the first proof of the fundamental theorem in the theory, a theorem that Borel had not believed to be generally true. However, the subject did not become important until 1944, when von Neumann and Morgenstern published their epoch-making book (**196**). (A second edition, with an appendix on cardinal utility measurement, came out in 1947 [**197**].) Their purpose was to analyze mathematically a very general class of problems, which might be called problems of strategy. Consider a game of tic-tac-toe. You know at any moment in the game what the moves available to your opponent are, but you do not know which one he will choose. The only information you have is that his choice will not, in general, be completely random; he will make a move which is designed in some way to increase his chance of winning and diminish yours. Thus the situation is one of uncertainty rather than risk. Your goals are similar to your opponent's. Your problem is: what strategy should you adopt? The theory of games offers no practical help in developing strategies, but it does offer rules about how to choose

among them. In the case of tic-tac-toe, these rules are trivial, since either player can force a draw. But in more complicated games of strategy, these rules may be useful. In particular, the theory of games may be helpful in analyzing proper strategy in games having random elements, like the shuffling of cards, or the throwing of dice. It should be noted that the concept of a game is an exceedingly general concept. A scientist in his laboratory may be considered to be playing a game against Nature. (Note, however, that we cannot expect Nature to try to defeat the scientist.) Negotiators in a labor dispute are playing a game against one another. Any situation in which money (or some valuable equivalent) may be gained as the result of a proper choice of strategy can be considered as a game.

To talk about game theory, a few technical terms are necessary. A *strategy* is a set of personal rules for playing the game. For each possible first move on your part, your opponent will have a possible set of responses. For each possible response by your opponent, you will have a set of responses, and so on through the game. A strategy is a list which specifies what your move will be for every conceivable previous set of moves of the particular game you are playing. Needless to say, only for the simplest games (e.g., matching pennies) does this concept of strategy have any empirical meaning.

Associated with strategies are *imputations*. An imputation is a set of payments made as a result of a game, one to each player. In general, different imputations will be associated with different sets of strategies, but for any given set of strategies there may be more than one imputation (in games involving coalitions).

Imputation X is said to *dominate*

putation Y if one or more of the players has separately greater gains or smaller losses) in X than in Y and can, by acting together (in the case of more than one player), enforce the occurrence of X, or of some other imputation at least as good. The relation of domination is not transitive.

A *solution* is a set of imputations, none of which dominates another, such that every imputation outside the solution is dominated by at least one imputation within the solution. Von Neumann and Morgenstern assert that the task of the theory of games is to find solutions. For any game, there may be one or more than one. One bad feature of the theory of games is that it frequently gives a large, or even infinite, number of solutions for a game.

The above definitions make clear that the only determiner of behavior in games, according to this theory, is the amounts of money which may be won or lost, or the expected amounts in games with random elements. The fun of playing, if any, is irrelevant.

The minimax loss principle. The notions of domination and of solution imply a new fundamental rule for decision making—a rule sharply different from the rule of maximizing utility or expected utility with which this paper has been concerned up to this section. This rule is the rule of minimizing the maximum loss, or, more briefly, *minimax loss*. In other words, the rule is to consider, for each possible strategy that you could adopt, what the worst possible outcome is, and then to select that strategy which would have the least ill-effects if the worst possible outcome happened. Another way of putting the same idea is to call it the principle of maximizing the minimum gain, or *maximin gain*. This rule makes considerable sense in two-person games

when you consider that the other player is out to get you, and so will do his best to make the worst possible outcome for you occur. If this rule is expressed geometrically, it asserts that the point you should seek is a saddle point. Just like a horse-back in a mountain pass (the best rule for crossing mountains is to minimize the maximum height, so explorers seek out such saddle-points).

Before we go any further, we need a few more definitions. Games may be among any number of players, but the simplest game is a *two-person game*, and it is this kind of game which has been most extensively and most successfully analyzed. Fundamentally, two kinds of payoff arrangements are possible. The simplest and most common is the one in which one player wins what the other player loses, or, more generally, the one for which the sum of all the payments made as a result of the game is zero. This is called a *zero-sum game*. In *nonzero-sum games*, analytical complexities arise. These can be diminished by assuming the existence of a fictitious extra player, who wins or loses enough to bring the sum of payments back to zero. Such a fictitious player cannot be assumed to have a strategy and cannot, of course, interact with any of the other players.

In zero-sum two-person games, what will happen? Each player, according to the theory, should pick his minimax strategy. But will this result in a stable solution? Not always. Sometimes the surface representing the possible outcomes of the game does not have a saddle-point. In this case, if player A chooses his minimax strategy, then player B will have an incentive not to use his own minimax strategy, because having found out his opponent's strategy, he can gain more by some other strategy. Thus the game has no solution.

Various resolutions of this problem are possible. Von Neumann and Morgenstern chose to introduce the notion of a *mixed strategy*, which is a probability distribution of two or more pure strategies. The fundamental theorem of the theory of games is that if both players in a zero-sum two-person game adopt mixed strategies which minimize the maximum *expected loss*, then the game will always have a saddle-point. Thus each person will get, in the long run, his expected loss, and will have no incentive to change his behavior even if he should discover what his opponent's mixed strategy is. Since A is already getting the minimum possible under the strategy he chose, any change in strategy by B will only increase A's payoff, and therefore cause B to gain less or lose more than he would by his own minimax strategy. The same is true of B.

Games involving more than two people introduce a new element—the possibility that two or more players will cooperate to beat the rest. Such a cooperative agreement is called a *coalition*, and it frequently involves *side-payments* among members of the coalition. The method of analysis for three-or-more-person games is to consider all possible coalitions and to solve the game for each coalition on the principles of a two-person game. This works fairly well for three-person games, but gets more complicated and less satisfactory for still more people.

This is the end of this exposition of the content of von Neumann and Morgenstern's book. It is of course impossible to condense a tremendous and difficult book into one page. The major points to be emphasized are these: the theory of games is not a model of how people actually play games (some game theorists will dis-

agree with this), nor is it likely to be of any practical use in telling you how to play a complicated game; the crux of the theory of games is the principle of choosing the strategy which minimizes the maximum expected financial loss; and the theory defines a solution of a game as a set of imputations which satisfies this principle for all players.

Assumptions. In their book von Neumann and Morgenstern say "We have . . . assumed that [utility] is numerical . . . substitutable and unrestrictedly transferable between the various players." (**197**, p. 604.) Game theorists disagree about what this and other similar sentences mean. One likely interpretation is that they assume utility to be linear with the physical value of money involved in a game and to be interpersonally comparable. The linear utility curves seem to be necessary for solving two-person games; the interpersonal comparability is used for the extension to *n* persons. Attempts are being made to develop solutions free of these assumptions (**176**).

Statistical decision functions. Von Neumann (**195**) first used the minimax principle in his first publication on game theory in 1928. Neyman and Pearson mentioned its applicability to statistical decision problems in 1933 (**144**). Wald (**198**), who prior to his recent death was the central figure in the statistical decision-function literature, first seriously applied the minimax principle to statistical problems in 1939. Apparently, all these uses of the principle were completely independent of one another.

After *Theory of Games and Economic Behavior* appeared in 1944, Wald (**198**) reformulated the problem of statistical decision making as one of playing a game against Nature.

The statistician must decide, on the basis of observations which cost something to make, between policies, each of which has a possible gain or loss. In some cases, all of these gains and losses and the cost of observing can be exactly calculated, as in industrial quality control. In other cases, as in theoretical research, it is necessary to make some assumption about the cost of being wrong and the gain of being right. At any rate, when they are put in this form, it is obvious that the ingredients of the problem of statistical decision making have a gamelike sound. Wald applied the minimax principle to them in a way essentially identical with game theory.

A very frequent criticism of the minimax approach to games against Nature is that Nature is not hostile, as is the opponent in a two-person game. Nature will not, in general, use a minimax strategy. For this reason, other principles of decision making have been suggested. The simple principle of maximizing expected utility (which is the essence of the Bayes's theorem [15, 198] solution of the problem) is not always applicable because, even though Nature is not hostile, she does not offer any way of assigning a probability to each possible outcome. In other words, statistical decision making is a problem of uncertainty, rather than of risk. Savage has suggested the principle of minimaxing *regret*, where regret is defined as the difference between the maximum which can be gained under any strategy given a certain state of the world and the amount gained under the strategy adopted. Savage believes (170, also personal communication) that neither von Neumann and Morgenstern nor Wald actually intended to propose the principle of minimaxing loss; they

confined their discussions to cases in which the concepts of minimax loss and minimax regret amount to the same thing. Other suggested principles are: maximizing the maximum expected gain, and maximizing some weighted average of the maximum and minimum expected gains (93). None of these principles commands general acceptance; each can be made to show peculiar consequences under some conditions (see 170).

Experimental games. The concepts of the theory of games suggest a new field of experimentation: How do people behave in game situations? Such experimentation would center on the development of strategies, particularly mixed strategies, and, in three-or-more-person games, on the development of coalitions and on the bargaining process. You should remember that the theory of games does not offer a mathematical model predicting the outcomes of such games (except in a few special cases); all it does is offer useful concepts and language for talking about them, and predict that certain outcomes will not occur.

A few minor experiments of this kind have been conducted by Flood, a mathematician, while he was at Rand Corporation. He usually used colleagues, many of whom were experts in game theory, and secretaries as subjects. The general design of his experiments was that a group of subjects were shown a group of desirable objects on a table, and told that they, as a group, could have the first object they removed from the table, and that they should decide among themselves which object to choose and how to allocate it. In the first experiment (64) the allocation problem did not arise because enough duplicate objects were provided so that each subject could have one of

the kind of object the group selected. The subjects were Harvard undergraduates, and the final selection was made by negotiation and voting. In the second experiment (65), in which the subjects were colleagues and secretaries, a long negotiation process eliminated some of the objects, but a time limit forced a selection by lot from among the rest. Further negotiations to solve the allocation problem were terminated by a secretary, who snatched the object, announced that it was hers, and then tried to sell it. No one was willing to buy, so the experiment terminated. Other experiments (66, 67) showed that coalitions sometimes form, that a sophisticated subject could blackmail the group for an extra side-payment by threatening to change his vote, and that the larcenous secretary, having succeeded once, had to be physically restrained in subsequent sessions to prevent more larceny. The general conclusion suggested by all these experiments is that even experts on game theory are less rational and more conventional than game theory might lead experimenters to expect.

Psychological comments. The most nutritive research problems in this area seem to be the social problems of how bargaining takes place. Flood's experiments left bargainers free and used physical objects, whose utilities probably vary widely from subject to subject, as stimuli to bargain over. This is naturalistic, but produces data too complex and too nonnumerical for easy analysis. A simpler situation in which the possible communications from one bargainer to another are limited (perhaps by means of an artificial vocabulary), in which the subjects do not see one another, and in which the object bargained over is simple, preferably being merely a sum of money, would be better. Physical isolation of one subject from another would make it possible to match each subject against a standard bargainer, the experimenter or a stooge, who bargains by a fixed set of rules that are unknown to the subject. Flood (personal communication) is conducting experiments of this sort. For three-or-more-person games, Asch's (16) technique of using a group consisting of only one real subject and all the rest stooges might well be used. It would be interesting, for instance, to see how the probability of a coalition between two players changes as the number and power of players united against them increase.

The theory of games is the area among those described in this paper in which the uncontrolled and casually planned "pilot experiment" is most likely to occur. Such experiments are at least as dangerous here as they are in the area of risky decision making. Flood's results suggest that it is especially important to use naive subjects and to use them only once, unless the effects of expertness and experience are the major concern of the experiment.

SUMMARY

For a long time, economists and others have been developing mathematical theories about how people make choices among desirable alternatives. These theories center on the notion of the subjective value, or utility, of the alternatives among which the decider must choose. They assume that people behave rationally, that is, that they have transitive preferences and that they choose in such a way as to maximize utility or expected utility.

The traditional theory of riskless choices, a straightforward theory of utility maximization, was challenged by the demonstration that the mathe-

matical tool of indifference curves made it possible to account for riskless choices without assuming that utility could be measured on an interval scale. The theory of riskless choices predicted from indifference curves has been worked out in detail. Experimental determination of indifference curves is possible, and has been attempted. But utility measured on an interval scale is necessary (though not sufficient) for welfare economics.

Attention was turned to risky choices by von Neumann and Morgenstern's demonstration that complete weak ordering of risky choices implies the existence of utility measurable on an interval scale. Mosteller and Nogee experimentally determined utility curves for money from gambling decisions, and used them to predict other gambling decisions. Edwards demonstrated the existence of preferences among probabilities in gambling situations, which complicates the experimental measurement of utility. Coombs developed a model

for utility and subjective probability measured on an ordered metric scale, and did some experiments to test implications of the model.

Economists have become worried about the assumption that choices are transitive. Experiments have shown that intransitive patterns of choice do occur, and so stochastic models have been developed which permit occasional intransitivities.

The theory of games presents an elaborate mathematical analysis of the problem of choosing from among alternative strategies in games of strategy. This paper summarizes the main concepts of this analysis. The theory of games has stimulated interest in experimental games, and a few bargaining experiments which can be thought of in game-theoretical terms have been performed.

All these topics represent a new and rich field for psychologists, in which a theoretical structure has already been elaborately worked out and in which many experiments need to be performed.

REFERENCES

1. ALCHIAN, A. The meaning of utility measurement. *Amer. econ. Rev.*, 1953, 43, 26–50.
2. ALLAIS, M. Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école americaine. *Colloque Internationale du Centre National de la Recherche scientifique*, 1952, No. 36.
3. ALLAIS, M. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école americaine. *Econometrica*, 1953, 21, 503–546.
4. ALLAIS, M. L'Extension des théories de l'equilibre economique général et du rendement social au cas du risque. *Econometrica*, 1953, 21, 269–290.
5. ALLAIS, M. La psychologie de l'homme rationnel devant le risque: La théorie et l'expérience. *J. soc. Statist.*, Paris, 1953, 94, 47–73.

6. ALLEN, R. G. D. The nature of indifference curves. *Rev. econ. Stud.*, 1933, 1, 110–121.
7. ALLEN, R. G. D. A note on the determinateness of the utility function. *Rev. econ. Stud.*, 1934, 2, 155–158.
8. ARMSTRONG, W. E. The determinateness of the utility function. *Econ. J.*, 1939, 49, 453–467.
9. ARMSTRONG, W. E. Uncertainty and the utility function. *Econ. J.*, 1948, 58, 1–10.
10. ARMSTRONG, W. E. A note on the theory of consumer's behavior. *Oxf. econ. Pap.*, 1950, 2, 119–122.
11. ARMSTRONG, W. E. Utility and the theory of welfare. *Oxf. econ. Pap.*, 1951, 3, 259–271.
12. ARROW, K. J. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, 1951, 19, 404–437.
13. ARROW, K. J. An extension of the basic

theorems of classical welfare economics. In J. Neyman (Ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability.* Berkeley: Univ. of Calif. Press, 1951. Pp. 507–532.

14. ARROW, K. J. *Social choice and individual values.* New York: Wiley, 1951.

15. ARROW, K. J., BLACKWELL, D., & GIRSHICK, M. A. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 1949, **17**, 213–244.

16. ASCH, S. E. *Social psychology.* New York: Prentice-Hall, 1952.

17. ATTNEAVE, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, **46**, 81–86.

18. BAUMOL, W. J. Community indifference. *Rev. econ. Stud.*, 1946, **14**, 44–48.

19. BAUMOL, W. J. The Neumann-Morgenstern utility index—an ordinalist view. *J. polit. Econ.*, 1951, **59**, 61–66.

20. BAUMOL, W. J. Discussion. *Amer. econ. Rev. Suppl.*, 1953, **43**, 415–416.

21. BERGSON (BURK), A. Reformulation of certain aspects of welfare economics. *Quart. J. Econ.*, 1938, **52**, 310–334.

22. BERNARDELLI, H. Note on the determinateness of the utility function. *Rev. econ. Stud.*, 1934, **2**, 69–75.

23. BERNARDELLI, H. The end of marginal utility theory? *Economica*, 1938, **5**, 192–212.

24. BERNARDELLI, H. A reply to Mr. Samuelson's note. *Economica*, 1939, **6**, 88–89.

25. BERNARDELLI, H. A rehabilitation of the classical theory of marginal utility. *Economica*, 1952, **19**, 254–268.

26. BERNOULLI, D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperiales Petropolitanae*, 1738, **5**, 175–192. (Trans. by L. Sommer in *Econometrica*, 1954, **22**, 23–36.)

27. BILODEAU, E. A. Statistical versus intuitive confidence. *Amer. J. Psychol.*, 1952, **65**, 271–277.

28. BISHOP, R. L. Consumer's surplus and cardinal utility. *Quart. J. Econ.*, 1943, **57**, 421–449.

29. BISHOP, R. L. Professor Knight and the theory of demand. *J. polit. Econ.*, 1946, **54**, 141–169.

30. BOHNERT, H. G. The logical structure of the utility concept. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision Processes.* New York: Wiley, in press.

31. BOREL, E. La théorie du jeu et les équations intégrales à noyau symétrique. *C. R. Acad. Sci., Paris*, 1921, **173**, 1304–1308. (Trans. by L. J. Savage in *Econometrica*, 1953, **21**, 97–100.)

32. BOREL, E. Sur les jeux où interviennent l'hasard et l'habilité des joueurs. In E. Borel, *Théorie des probabilités.* Paris: Librairie Scientifique, J. Hermann, 1924. Pp. 204–224. (Trans. by L. J. Savage in *Econometrica*, 1953, **21**, 101–115.)

33. BOREL, E. Algèbre et calcul des probabilités. *C. R. Acad. Sci., Paris*, 1927, **184**, 52–53. (Trans. by L. J. Savage in *Econometrica*, 1953, **21**, 116–117.)

34. BOREL, E. *Traité du calcul des probabilités et de ses applications, applications des jeux de hasard.* Vol. IV, No. 2. Paris: Gauthier-Villars, 1938.

35. BROSS, I. *Design for decision.* New York: Macmillan, 1953.

36. BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.

37. BUSH, R. R., & MOSTELLER, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.

38. CARTWRIGHT, D. Decision-time in relation to differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.

39. CARTWRIGHT, D. The relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196.

40. CARTWRIGHT, D. Survey research: psychological economics. In J. G. Miller (Ed.), *Experiments in social process.* New York: McGraw-Hill, 1950. Pp. 47–64.

41. CARTWRIGHT, D., & FESTINGER, L. A quantitative theory of decision. *Psychol. Rev.*, 1943, **50**, 595–621.

42. CLARK, J. M. Realism and relevance in the theory of demand. *J. polit. Econ.*, 1946, **54**, 347–353.

43. COOMBS, C. H. Psychological scaling without a unit of measurement. *Psychol. Rev.*, 1950, **57**, 145–158.

44. COOMBS, C. H. Mathematical models in psychological scaling. *J. Amer. statist. Ass.*, 1951, **46**, 480–489.

45. COOMBS, C. H. A theory of psychological scaling. *Bull. Engng Res. Inst. Univer. Mich.*, 1952, No. 34.

46. COOMBS, C. H. Theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in*

the behavioral sciences. New York: Dryden, 1953. Pp. 471–535.

47. COOMBS, C. H. A method for the study of interstimulus similarity. *Psychometrika*, in press.

48. COOMBS, C. H., & BEARDSLEE, D. C. Decision making under uncertainty. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, in press.

49. COOMBS, C. H., & MILHOLLAND, J. E. Testing the "rationality" of an individual's decision making under uncertainty. *Psychometrika*, in press.

50. CORLETT, W. J., & NEWMAN, P. K. A note on revealed preference and the transitivity conditions. *Rev. econ. Stud.*, 1952, **20**, 156–158.

51. DE FINETTI, B. La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. Poincaré*, 1937, **7**, 1–68.

52. DE FINETTI, B. Recent suggestions for the reconciliation of theories of probability. In J. Neyman (Ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability.* Berkeley: Univer. of Calif. Press, 1951.

53. EDGEWORTH, F. Y. *Mathematical psychics.* London: Kegan Paul, 1881.

54. EDWARDS, W. Discussion. *Econometrica*, 1953, **21**, 477. (Abstract)

55. EDWARDS, W. Experiments on economic decision-making in gambling situations. *Econometrica*, 1953, **21**, 349–350. (Abstract)

56. EDWARDS, W. Information, repetition, and reinforcement as determiners of two-alternative decisions. *Amer. Psychologist*, 1953, **8**, 345. (Abstract)

57. EDWARDS, W. Probability-preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

58. EDWARDS, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67.

59. EDWARDS, W. The reliability of probability preferences. *Amer. J. Psychol.*, 1954, **67**, 68–95.

60. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

61. FESTINGER, L. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus differences. *J. exp. Psychol.*, 1943, **32**, 291–306.

62. FESTINGER, L. Studies in decision: II.

An empirical test of a quantitative theory of decision. *J. exp. Psychol.*, 1943, **32**, 411–423.

63. FISHER, I. A statistical method for measuring "marginal utility" and testing the justice of a progressive income tax. In J. Hollander (Ed.), *Economic essays contributed in honor of John Bates Clark.* New York: Macmillan, 1927. Pp. 157–193.

64. FLOOD, M. M. A preference experiment. *Rand Corp. Memo.*, November 1951, No. P-256.

65. FLOOD, M. M. A preference experiment (Series 2, Trial 1). *Rand Corp. Memo.*, December 1951, No. P-258.

66. FLOOD, M. M. A preference experiment (Series 2, Trials 2, 3, 4). *Rand Corp. Memo.*, January 1952, No. P-263.

67. FLOOD, M. M. A preference experiment (Series 3). Unpublished memorandum, Rand Corporation. February 25, 1952.

68. FLOOD, M. M. Some experimental games. *Rand Corp. Memo.*, March 1952, No. RM-789-1. (Revised June 1952.)

69. FLOOD, M. M. Testing organization theories. *Rand Corp. Memo.*, November 1952, No. P-312.

70. FLOOD, M. M. An experimental multiple-choice situation. *Rand Corp. Memo.*, November 1952, No. P-313.

71. FRÉCHET, M. Emile Borel, initiator of the theory of psychological games and its application. *Econometrica*, 1953, **21**, 95–96.

72. FRÉCHET, M., & VON NEUMANN, J. Commentary on the three notes of Emile Borel. *Econometrica*, 1953, **21**, 118–126.

73. FRIEDMAN, M., & SAVAGE, L. J. The utility analysis of choices involving risk. *J. polit. Econ.*, 1948, **56**, 279–304. (Reprinted with minor changes in G. J. Stigler & K. E. Boulding [Eds.], *Readings in price theory.* Chicago: Richard D. Irwin, 1952. Pp. 57–96.)

74. FRIEDMAN, M., & SAVAGE, L. J. The expected-utility hypothesis and the measurability of utility. *J. polit. Econ.*, 1952, **60**, 463–475.

75. FRISCH, R. New methods of measuring marginal utility. In R. Frisch, *Beiträge zur ökonomischen Theorie.* Tübingen: Mohr, 1932.

76. GEORGESCU-ROEGEN, N. The pure theory of consumer's behavior. *Quart. J. Econ.*, 1936, **50**, 545–593.

77. GEORGESCU-ROEGEN, N. The theory of

choice and the constancy of economic laws. *Quart. J. Econ.*, 1950, 64, 125–138.

78. GEORGESCU-ROEGEN, N. Utility, expectations, measurability, prediction. Paper read at Econometric Soc., Kingston, September, 1953.

79. GIRSHICK, M. A., & BLACKWELL, D. *Theory of games and statistical decisions.* New York: Wiley, 1954.

80. GOOD, I. J. *Probability and the weighing of evidence.* London: Griffin, 1950.

81. GRIFFITH, R. M. Odds adjustments by American horse-race bettors. *Amer. J. Psychol.*, 1949, 62, 290–294.

82. HARSANYI, J. C. Cardinal utility in welfare economics and in the theory of risk-taking. *J. polit. Econ.*, 1953, 61, 434–435.

83. HART, A. G. Risk, uncertainty, and the unprofitability of compounding probabilities. In O. Lange, F. McIntyre, & T. O. Yntema (Eds.), *Studies in mathematical economics and econometrics.* Chicago: Univer. of Chicago Press, 1942. Pp. 110–118.

84. HAYES, S. P., JR. Some psychological problems of economics. *Psychol. Bull.*, 1950, 47, 289–330.

85. HERSTEIN, I. N., & MILNOR, J. An axiomatic approach to measurable utility. *Econometrica*, 1953, 21, 291–297.

86. HICKS, J. R. The foundations of welfare economics. *Econ. J.*, 1939, 49, 696–712.

87. HICKS, J. R. *Value and capital.* Oxford: Clarendon Press, 1939.

88. HICKS, J. R., & ALLEN, R. G. D. A reconsideration of the theory of value. *Economica*, 1934, 14, 52–76, 196–219.

89. HILDRETH, C. Alternative conditions for social orderings. *Econometrica*, 1953, 21, 81–94.

90. HOUTHAKKER, H. S. Revealed preference and the utility function. *Economica*, 1950, 17, 159–174.

91. HULL, C. L. *Principles of behavior, an introduction to behavior theory.* New York: D. Appleton-Century, 1943.

92. HURWICZ, L. The theory of economic behavior. *Amer. econ. Rev.*, 1945, 35, 909–925. (Reprinted in G. J. Stigler & K. E. Boulding [Eds.], *Readings in price theory.* Chicago: Richard D. Irwin, 1952. Pp. 505–526.)

93. HURWICZ, L. What has happened to the theory of games? *Amer. econ. Rev. Suppl.*, 1953, 43, 398–405.

94. IRWIN, F. W. Stated expectations as

functions of probability and desirability of outcomes. *J. Pers.*, 1953, 21, 329–335.

95. JARRETT, JACQUELINE M. Strategies in risk-taking situations. Unpublished Ph.D. thesis, Harvard Univer., 1951.

96. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, 47, 193–234.

97. JOHNSON, W. E. The pure theory of utility curves. *Econ. J.*, 1913, 23, 483–513.

98. KALDOR, N. Welfare propositions and inter-personal comparisons of utility. *Econ. J.*, 1939, 49, 549–552.

99. KALDOR, N. A comment. *Rev. econ. Stud.*, 1946, 14, 49.

100. KAPLAN, A., & RADNER, R. A questionnaire approach to subjective probability—some experimental results. Working Memorandum 41, Santa Monica Conference on Decision Problems, August 15, 1952.

101. KATONA, G. Psychological analysis of business decisions and expectations. *Amer. econ. Rev.*, 1946, 36, 44–62.

102. KATONA, G. Contributions of psychological data to economic analysis. *J. Amer. statist. Ass.*, 1947, 42, 449–459.

103. KATONA, G. *Psychological analysis of economic behavior.* New York: McGraw-Hill, 1951.

104. KATONA, G. Rational behavior and economic behavior. *Psychol. Rev.*, 1953, 60, 307–318.

105. KAUDER, E. Genesis of the marginal utility theory from Aristotle to the end of the eighteenth century. *Econ. J.*, 1953, 63, 638–650.

106. KAUDER, E. The retarded acceptance of the marginal utility theory. *Quart. J. Econ.*, 1953, 67, 564–575.

107. KAYSEN, C. A revolution in economic theory? *Rev. econ. Stud.*, 1946, 14, 1–15.

108. KENNEDY, C. The common sense of indifference curves. *Oxf. econ. Pap.*, 1950, 2, 123–131.

109. KNIGHT, F. H. *Risk, uncertainty, and profit.* Boston: Houghton Mifflin, 1921.

110. KNIGHT, F. H. Realism and relevance in the theory of demand. *J. polit. Econ.*, 1944, 52, 289–318.

111. KNIGHT, F. H. Comment on Mr. Bishop's article. *J. polit. Econ.*, 1946, 54, 170–176.

112. KUHN, H. W., & TUCKER, A. W. (Eds.) Contributions to the theory of games.

Vol. I. *Ann. Math. Stud.*, No. 24. Princeton: Princeton Univer. Press, 1950.

113. KUHN, H. W., & TUCKER, A. W. (Eds.) Contributions to the theory of games. Vol. II. *Ann. Math. Stud.*, No. 28. Princeton: Princeton Univer. Press, 1953.

114. LANCASTER, K. A refutation of Mr. Bernardelli. *Economica*, 1953, 20, 259–262.

115. LANDAHL, H. D. A neurobiophysical interpretation of certain aspects of the problem of risks. *Bull. Math. Biophysics*, 1951, 13, 323–335.

116. LANGE, O. The determinateness of the utility function. *Rev. econ. Stud.*, 1933, 1, 218–225.

117. LANGE, O. Note on the determinateness of the utility function. *Rev. econ. Stud.*, 1934, 2, 75–77.

118. LANGE, O. The foundations of welfare economics. *Econometrica*, 1942, 10, 215–228.

119. LANGE, O. The scope and methods of economics. *Rev. econ. Stud.*, 1945, 13, 19–32.

120. LEWIN, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936.

121. LEWIN, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 791–844.

122. LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. I. New York: Ronald, 1944. Pp. 333–378.

123. LEWISOHN, S. A. Psychology in economics. *Polit. Sci. Quart.*, 1938, 53, 233–238.

124. LITTLE, I. M. D. The foundations of welfare economics. *Oxf. econ. Pap.*, 1949, 1, 227–246.

125. LITTLE, I. M. D. A reformulation of the theory of consumer's behavior. *Oxf. econ. Pap.*, 1949, 1, 90–99.

126. LITTLE, I. M. D. The theory of consumer's behavior—a comment. *Oxf. econ. Pap.*, 1950, 2, 132–135.

127. LITTLE, I. M. D. Social choice and individual values. *J. polit. Econ.*, 1952, 60, 422–432.

128. MACFIE, A. L. Choice in psychology and as economic assumption. *Econ. J.*, 1953, 63, 352–367.

129. McKINSEY, J. C. C. *Introduction to the*

theory of games. New York: McGraw-Hill, 1952.

130. MALINVAUD, E. Note on von Neumann-Morgenstern's strong independence axiom. *Econometrica*, 1952, 20, 679.

131. MANNE, A. S. The strong independence assumption—gasolene blends and probability mixtures. *Econometrica*, 1952, 20, 665–669.

132. MARKOWITZ, H. The utility of wealth. *J. polit. Econ.*, 1952, 60, 151–158.

133. MARKS, ROSE W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, 19, 332–351.

134. MARSCHAK, J. Neumann's and Morgenstern's new approach to static economics. *J. polit. Econ.*, 1946, 54, 97–115.

135. MARSCHAK, J. Rational behavior, uncertain prospects, and measurable utility. *Econometrica*, 1950, 18, 111–141.

136. MARSCHAK, J. Why "should" statisticians and businessmen maximize "moral expectation"? In J. Neyman (Ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability.* Berkeley: Univer. of Calif. Press, 1951. Pp. 493–506.

137. MARSHALL, A. *Principles of economics.* (8th Ed.) New York: Macmillan, 1948.

138. MAY, K. O. Transitivity, utility, and aggregation in preference patterns. *Econometrica*, 1954, 22, 1–13.

139. MELVILLE, L. G. Economic welfare. *Econ. J.*, 1939, 49, 552–553.

140. MISHAN, E. J. The principle of compensation reconsidered. *J. polit. Econ.*, 1952, 60, 312–322.

141. MORGAN, J. N. Can we measure the marginal utility of money? *Econometrica*, 1945, 13, 129–152.

142. MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, 59, 371–404.

143. NEISSER, H. The strategy of expecting the worst. *Soc. Res.*, 1952, 19, 346–363.

144. NEYMAN, J., & PEARSON, E. S. The testing of statistical hypotheses in relation to probability *a priori. Proc. Cambr. phil. Soc.*, 1933, 29, 492–510.

145. PAPANDREOU, A. G. An experimental test of an axiom in the theory of choice. *Econometrica*, 1953, 21, 477. (Abstract)

146. PARETO, V. *Manuale di economia politica, con una introdusione ulla scienza*

sociale. Milan, Italy: Societa Editrice Libraria, 1906.

147. PHELPS-BROWN, E. H. Note on the determinateness of the utility function. *Rev. econ. Stud.*, 1934, 2, 66–69.

148. PIGOU, A. C. Some aspects of welfare economics. *Amer. econ. Rev.*, 1951, 41, 287–302.

149. PRESTON, M. G., & BARATTA, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, 61, 183–193.

150. RAMSEY, F. P. Truth and probability. In F. P. Ramsey, *The foundations of mathematics and other logical essays.* New York: Harcourt Brace, 1931.

151. RICCI, U. Pareto and pure economics. *Rev. econ. Stud.*, 1933, 1, 3–21.

152. ROBBINS, L. Interpersonal comparisons of utility: a comment. *Econ. J.*, 1938, 48, 635–641.

153. ROBBINS, L. Robertson on utility and scope. *Economica*, 1953, 20, 99–111.

154. ROBERTSON, D. H. *Utility and all that and other essays.* London: George Allen & Unwin, 1952.

155. ROTHENBERG, J. Conditions for a social welfare function. *J. polit. Econ.*, 1953, 61, 389–405.

156. ROTHSCHILD, K. W. The meaning of rationality: a note on Professor Lange's article. *Rev. econ. Stud.*, 1946, 14, 50–52.

157. ROUSSEAS, S. W., & HART, A. G. Experimental verification of a composite indifference map. *J. polit. Econ.*, 1951, 59, 288–318.

158. SAMUELSON, P. A. A note on measurement of utility. *Rev. econ. Stud.*, 1937, 4, 155–161.

159. SAMUELSON, P. A. Empirical implications of utility analysis. *Econometrica*, 1938, 6, 344–356.

160. SAMUELSON, P. A. A note on the pure theory of consumer's behavior. *Economica*, 1938, 5, 61–71.

161. SAMUELSON, P. A. A note on the pure theory of consumer's behavior. An addendum. *Economica*, 1938, 5, 353–354.

162. SAMUELSON, P. A. The numerical representations of ordered classifications and the concept of utility. *Rev. econ. Stud.*, 1938, 6, 65–70.

163. SAMUELSON, P. A. The end of marginal utility: a note on Dr. Bernardelli's article. *Economica*, 1939, 6, 86–87.

164. SAMUELSON, P. A. *Foundations of economic analysis.* Cambridge, Mass.: Harvard Univer. Press, 1947.

165. SAMUELSON, P. A. Consumption theory in terms of revealed preference. *Economica*, 1948, 15, 243–253.

166. SAMUELSON, P. A. Evaluation of real national income. *Oxf. econ. Pap.*, 1950, 2, 1–29.

167. SAMUELSON, P. A. The problem of integrability in utility theory. *Economica*, 1950, 17, 355–385.

168. SAMUELSON, P. A. Probability, utility, and the independence axiom. *Econometrica*, 1952, 20, 670–678.

169. SAMUELSON, P. A. Consumption theorems in terms of overcompensation rather than indifference comparisons. *Economica*, 1953, 20, 1–9.

170. SAVAGE, L. J. The theory of statistical decision. *J. Amer. statist. Ass.*, 1951, 46, 55–67.

171. SAVAGE, L. J. An axiomatic theory of reasonable behavior in the face of uncertainty. Unpublished manuscript, Statistical Research Center, Univer. of Chicago, No. SRC-21222S14.

172. SCHULTZ, H. *The theory and measurement of demand.* Chicago: Univer. of Chicago Press, 1938.

173. SCITOVSKY, T. A note on welfare propositions in economics. *Rev. econ. Stud.*, 1941, 9, 77–88.

174. SCITOVSKY, T. The state of welfare economics. *Amer. econ. Rev.*, 1951, 41, 303–315.

175. SHACKLE, G. L. S. *Expectations in economics.* Cambridge, Eng.: Cambridge Univer. Press, 1949.

176. SHAPLEY, L. S., & SHUBIK, M. Solutions of n-person games with ordinal utilities. *Econometrica*, 1953, 21, 348–349 (Abstract)

177. SLUTSKY, E. E. Sulla teoria del bilancio del consumatore, *Giornale degli economisti*, 1915, 51, 1–26. (Trans. by O. Ragusa and reprinted in G. J. Stigler & K. E. Boulding [Eds.], *Readings in price theory.* Chicago: Richard D. Irwin, 1952. Pp. 27–56.)

178. SPROWLS, R. C. Psychological-mathematical probability in relationships of lottery gambles. *Amer. J. Psychol.*, 1953, 66, 126–130.

179. STIGLER, G. J. The limitations of statistical demand curves. *J. Amer. statist. Ass.*, 1939, 34, 469–481.

180. STIGLER, G. J. The development of utility theory. *J. polit. Econ.*, 1950, 58, 307–327, 373–396.

181. STONE, J. R. N. The theory of games. *Econ. J.*, 1948, 58, 185–201.

182. STONE, R. (J. R. N.) *The role of measure-*

ment in economics. Cambridge, Eng.: Cambridge Univer. Press, 1951.

183 STROTZ, R. H. Cardinal utility. *Amer. econ. Rev. Suppl.*, 1953, 43, 384–405.

184 SWEEZY, A. R. The interpretation of subjective value theory in the writings of the Austrian economists. *Rev. econ. Stud.*, 1933, 1, 176–185.

185 THURSTONE, L. L. The indifference function. *J. soc. Psychol.*, 1931, 2, 139–167.

186 THURSTONE, L. L. The measurement of values. *Psychol. Rev.*, 1954, 61, 47–58.

187 TINTNER, G. The theory of choice under subjective risk and uncertainty. *Econometrica*, 1941, 9, 298–304.

188 TINTNER, G. A contribution to the non-static theory of choice. *Quart. J. Econ.*, 1942, 56, 274–306.

189 TÖRNQVIST, L. On the economic theory of lottery-gambles. *Skand. Aktuar-Tidskr.*, 1945, 28, 228–246.

190 VAIL, S. V. Expectations, degrees of belief, psychological probabilities. Unpublished manuscript Univer. of Michigan, Seminar on the Application of Mathematics to the Social Sciences, October 23, 1952.

191 VAIL, S. V. A stochastic model of utilities. Unpublished manuscript, No. 24, Univer. of Michigan, Seminar on the Applications of Mathematics to the Social Sciences, April 23, 1953.

192 VAIL, S. V. Alternative calculi of subjective probabilities. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, in press.

193 VICKREY, W. S. Measuring marginal utility by reactions to risk. *Econometrica*, 1945, 13, 319–333.

194 VINER, J. The utility concept in value theory and its critics. *J. polit. Econ.*, 1925, 33, 369–387, 638–659.

195 VON NEUMANN, J. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 1928, 100, 295–320.

196 VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior.* (1st Ed.) Princeton: Princeton Univer. Press, 1944.

197 VON NEUMANN, J., & MORGENSTERN, O. *Theory of games and economic behavior.* (2nd Ed.) Princeton: Princeton Univer. Press, 1947.

198 WALD, A. *Statistical decision functions.* New York: Wiley, 1950.

199 WALKER, K. F. The psychological assumptions of economics. *Econ. Rec.*, 1946, 22, 66–82

200 WALLIS, W. A., & FRIEDMAN, M. The empirical derivation of indifference functions. In O. Lange, F. McIntyre, & T. O. Yntema, (Eds.), *Studies in mathematical economics and econometrics.* Chicago: Univer. of Chicago Press, 1942.

201 WECKSTEIN, R. S. On the use of the theory of probability in economics. *Rev. econ. Stud.*, 1953, 20, 191–198.

202 WEISSKOPF, W. A. Psychological aspects of economic thought. *J. polit. Econ.*, 1949, 57, 304–314.

203 WELDON, J. C. A note on measures of utility. *Canad. J. Econ. polit. Sci.*, 1950, 16, 227–233

204 WOLD, H. A synthesis of pure demand analysis. Part I. *Skand. Aktuar-Tidskr.*, 1943, 26, 85–118.

205 WOLD, H. A synthesis of pure demand analysis. Part II. *Skand. Aktuar-Tidskr.*, 1943, 26, 220–263.

206 WOLD, H. A synthesis of pure demand analysis. Part III. *Skand. Aktuar-Tidskr.*, 1944, 27, 69–120.

207 WOLD, H. Ordinal preferences or cardinal utility? *Econometrica*, 1952, 20, 661–664.

208 WOLD, H., & JURÉEN, L. *Demand analysis.* New York: Wiley, 1953.

209 ZEUTHEN, F. On the determinateness of the utility function. *Rev. econ. Stud.*, 1937, 4, 236–239.

SPECIAL REVIEW

SEXUAL BEHAVIOR IN THE HUMAN FEMALE[1,2]

HERBERT HYMAN[3] AND JOSEPH E. BARMACK
Columbia University *College of the City of New York*

PART I

THE SOCIAL PSYCHOLOGICAL DATA

Five years after the report on men, the publication of *Sexual Behavior in the Human Female* by Kinsey and his collaborators provides the companion results on women. It provides the greatest wealth of data ever accumulated on the sexual behavior of a particular population, American women, and therefore is a factual contribution of great interest to psychologists. To social psychologists and those interested in the relations of man to society, the new volume underscores and rounds out a basic finding of the first volume—namely, the widespread occurrence of a variety of unsanctioned forms of sexual behavior both among males and females, and the disparity between societal norms and such behavior.

While the novelty of this finding may have worn off slightly, the new volume has other compensating fea-

[1] KINSEY, A. C., POMEROY, W. B., MARTIN, C. E., & GEBHARD, P. H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953. Pp. xxx+842. $8.00.

[2] This multiple review was divided so that H. Hyman was concerned with Part I, the social psychological data, and J. E. Barmack confined himself to Part II, the biological material. The reviewers worked independently of each other.

[3] The reviewer wishes to acknowledge his indebtedness to Paul B. Sheatsley, with whom he collaborated on a longer essay on the Kinsey reports (cf. GEDDES, D. P. [Ed.] *An analysis of the Kinsey reports on sexual behavior in the human male and female.* New York: New American Library of World Literature, and E. P. Dutton, 1954).

tures. The simple fact that it is the *second* report suggests that the researchers may well have benefited from their past experience, with consequent improvements in the quality of their findings. This not only provides opportunity for improvement of methods, but it also permits formerly neglected areas of inquiry to be covered. Thus, in the earlier volume the emotional and attitudinal accompaniments of various patterns of behavior were hardly touched. We shall note below that considerable information of this type has been added, making the volume more relevant to psychologists.

In addition, the new volume is not merely a descriptive study like the first; Kinsey now can *compare* the behavior of males and females. The comparative profiles of the sexual behavior of men and women provide evidence on the degree of "complementarity" of partners. It is such data that are relevant to profound questions of conflict in marital and interpersonal relations, rather than separate descriptive data of the behavior of one group or the other. And much of the new volume exploits these materials; each chapter ends with a comparison of men and women, and a special chapter is devoted to general psychological differences between the sexes. Further, by parallel analyses of the *determinants* of sexual patterns in men and women, there emerge data on the *differential* and lesser impact of social factors on the sexual development of women—a particularly relevant body of materials for

studies of socialization. Parallel to these findings are other new materials relevant to general problems of learning theory and psychological development. In both volumes, sexual behavior was examined in relation to various group memberships. But now, supplementary analyses of a more dynamic sort are introduced. Kinsey studies the relation of the individual's early sexual patterns to his subsequent sexual behavior and adjustment. Obviously, such information is valuable both to clinicians and theorists.

The substantive contributions outlined above can make the new volume of great importance, provided the methods employed were such as to yield sound findings. It is to the question of methods therefore that this review will be directed.

This question cannot be answered briefly or in any simple way. The official committee appointed by the American Statistical Association to review the statistical methods employed in the *first* Kinsey report did not issue its report until three years after it had been established as a committee. With such distinguished members as W. G. Cochran, Frederick Mosteller, and John W. Tukey, it not only examined the published work, but it also spent a period of time at the Institute at Indiana University and engaged in certain additional researches and examination of other literature. The detailed evaluation will require monographic publication, and the summary findings will require some 43 pages of journal treatment. It is of note that this expert, judicial, and reasonable body remarks:

It would have been possible to write two factually correct reports, one of which would leave the impression with the reader that KPM's work was of the highest quality, the

other that the work was of poor quality and that the major issues were evaded. We have not written either of these extreme reports (1, p. 674).

Evaluation of the methods will require a brief discussion of each stage in the new inquiry: the nature of the sample of 5,940 white American females who provide the major data; the quality of the interviewing; the adequacy of the coverage of the interview schedule and its construction; and analysis of the data. We must also allude to the methods in the earlier volume. Since much of the significance of the new work derives from the comparative data on male and female, one must examine these data in an attempt to discover whether or not the differences demonstrated are an artifact of some change in method.

One reasonable standard to apply to this volume is the level of improvement in methods, as compared with the earlier report. The novelty of the earlier investigation was bound to create error. But the many criticisms of the first volume, and the additional experience, should have led to considerable improvement. That this is the case seems beyond question. Certain major defects have been remedied. Thus, there is much greater attention to error and more explicit reference to limitations. However, it should be noted that there are limits to the amount and kinds of improvement that could be achieved in the second volume. While the publication date may suggest that the new work had the benefit of some years of additional experience, this is clearly not the case. Both inquiries started in 1938 and "throughout the years, female histories have been added at approximately the same rate as the histories of males." Thus, Kinsey's sampling

and interviewing procedures cannot have been altered appreciably. What is subject to change are the analytic procedures since, in the processing and reporting of female data, Kinsey had the benefit of additional time. However, even here changes cannot be made lightly. For the comparative findings on males and females to be valid, the methods must be comparable. Bearing in mind such limitations, we can clearly see that Kinsey has made considerable improvement.

The sample. Critics of the first volume questioned Kinsey's generalizations to the whole American male population from his sample data. While they varied with respect to the inadmissibility of certain *procedures*, there was much agreement that the generalizations departed considerably from proper caution. In the present volume, Kinsey is cautious about generalizations, and he makes no attempt to extend his findings to all American women. A telling criticism of the earlier report was that no reader could determine just how representative of any particular universe the sample was, because nowhere was there any systematic account of the distribution of the 5,300 males in terms of standard characteristics. This defect has now been remedied by means of two elaborate tables which present detailed information on the composition of the 5,940 white females.

While Kinsey has this time provided us with these two tables and has carefully refrained from making unsupported generalizations, it can be demonstrated that the sample does not accurately represent the white female population of the nation. Some critics may condemn the whole investigation, arguing that a better method could have been employed—

one which would have permitted unqualified statements about the population as a whole. The question is debatable. Kinsey argues that any scientific sampling of predesignated respondents would have been impossible because a high proportion of such selected individuals could not be counted upon to consent to the interview and to answer the questions truthfully. In the present volume, all respondents were "volunteers" in the sense that they allowed themselves to be interviewed. However, histories were not accepted simply because a subject expressed the desire to be interviewed; all the respondents were selected because of their importance to the over-all sampling plan. Kinsey works primarily through formal or informal groups and makes ingenious and skilful use of social pressures. He describes some group contacts which required a year or two of cultivation. Indeed, the immense diversity of groups he has worked with is a tribute to his persistence and skill.

Nevertheless, one must still raise grave questions concerning the sample. For one thing, a substantial portion of the U. S. population claims no organizational membership whatever. Such individuals would likely be excluded from the samples, and there is evidence that they differ in many respects, perhaps including sexual behavior, from those who do belong to groups and associations. In addition, Kinsey could seldom interview all or even nearly all of the members of any particular group. No figures are given on the proportion of "refusals" to be interviewed, though Kinsey does state that 15% of the female sample is composed of "100% groups"—that is, groups in which every member was interviewed—and "we may report that a consider-

ble proportion of the rest of the sample has been drawn from groups in which something between 50 and 90 per cent of all the members had contributed histories" (p. 30). Kinsey further states: "Such coverage should provide a good sample of those particular groups" (p. 30), but this is a dubious assumption. There is no reason to believe that the histories of the 10–50% who do not contribute would necessarily agree with those who do. A principal finding of the Kinsey reports (to many, their most controversial and significant finding) has been the high incidences and frequencies of both sanctioned and unsanctioned forms of sexual behavior reported by both men and women. A speculation which cannot be answered concerns the possibility that those women who agree to an interview might be less inhibited and consequently may engage in more different types of sexual activity, and may do so more frequently.

Aside from this consideration, it is puzzling that Kinsey's sample is not more nearly representative of the population in respect to a number of gross characteristics. Perhaps the most striking deficiency is the failure to interview enough females whose education did not extend beyond grammar school. Some of the most interesting differences in the sexual behavior of the males studied were those attributed to education, and one would have expected comparative data in the present study. Instead, only 3% of the women interviewed had not attended high school, and Kinsey is restrained from making any substantial statements about this important stratum of the population. On the other hand, 75% of the total female sample had attended college, and 19% had gone on to

postgraduate work. The significance of Kinsey's findings must be considerably diluted when we consider that three-fourths of his sample came from the 13% of American women who have gone to college, and that the 40% of American women who never went beyond the eighth grade comprised only 3% of those he studied. Similarly, the sample is deficient in the older age groups and is heavily weighted with women in their 30's. Some of these sampling biases have been deliberate in order to build up sufficient cases among particular subgroups for special analyses. Thus, although Jews represent only about 4% of the U. S. population, they account for more than one-quarter of Kinsey's sample. In order to make statements about religion and sexual behavior, Kinsey did require sufficient Jews to allow thorough analysis of this religious group. In the case of the age and education biases, however, this reason does not seem relevant, and one can only wonder at the failure to achieve more adequate representation in terms of these two important factors.

Obviously one should not dismiss Kinsey's study simply because he cannot speak authoritatively about every group in the population. However, one feature of the present volume does raise a *new* question about the sampling method; that is the comparative analysis of male and female behavior. Many statistical data demonstrating sex differences are presented, and their impact is powerful. But a problem that continually arises in the interpretation of these findings, and one that is *not* stressed by Kinsey, is whether the differences represent true sex differences or merely *differential* sampling biases in the two sets of data.

If the biases were equivalent for both men and women, the stated differences are nonetheless true, though limited in generality. But if the male sample is biased in one respect (e.g., too many uneducated) and the females in an opposite respect (too many educated), then the interpretation of any differences found in the data becomes very difficult. Since the over-all composition of the male sample was never reported in the first volume, the facts cannot be readily established. Attempts to infer this composition from scattered tables suggest that some of the sex differences may be artifacts of differential sampling biases. Thus, almost 30% of the women are Jewish, while the best guess for the men would place only about 15% in that class. Only 3% of the women are in the grammar school group, but in the male sample this proportion is much larger. If religious affiliation and education are significantly correlated with both male and female sexual behavior, the differential biases in the two samples could well explain some of the aggregate sex differences reported.

However, the criticism is not as serious as it might seem. Kinsey presents much evidence that female sexual behavior is less affected by many of the social characteristics studied than is male behavior. Thus, biases with respect to these factors could not account for very much of the difference reported. Furthermore, Kinsey presents in summary form in the female volume, and in detail in the two separate volumes, the male and female findings for various subgroups, so that the careful reader can estimate the degree to which such biases may affect the overall differences cited. Nevertheless, any differential biases with respect to

characteristics other than those ... lated might still remain as an ... known source of error in the ... parisons. Thus, Kinsey states that female sexual behavior is subject to much greater *variability* than the male. Since some other unknown characteristics must determine the variation among females, any such differential sampling biases might well contribute to a spurious finding. Kinsey does present some empirical evidence bearing on the gener... problem when, for example, he shows a close agreement between the male and female findings with respect to marital coitus. The frequencies of marital coitus would, by definition have to be identical in the aggregate for the two samples if the samples were equivalent. In many instances, a very high degree of agreement is reported, but for some few aspects of behavior the reports of the two samples do not closely conform. The speculation is not *generally* warranted, but some limited number of the reported sex differences may well reflect the net operation of all the differential biases in the sampling of the two groups.

The interview schedule and the interviewers. As in the case of the males, approximately two hours were spent interviewing each subject and covering between 300 and 500 items of information. No questionnaire was presented in either volume. This effectively prevents one from appraising the types of questions that Kinsey asked and the manner in which they were presented. Kinsey argues persuasively against the belief that "standard questions fed through diverse human machines can bring standard answers," and he used no questionnaire at all in the usual sense of the word. The items to be covered and the definition of each

item were standard, but the wording and order of the particular questions were varied in the most meaningful manner for each respondent. Since the objective conditions of the interview were not uniform, one can never be certain that the differences found between individuals and groups are not due, at least in part, to differences in the wording and order of the questions. This consideration takes on added weight in connection with the many comparisons between males and females. Here, the question of consistency of interviewing procedure might be raised, but unfortunately can never be answered.

Offsetting this possible danger was the fact that all of the interviewing was done by a very *small* staff of highly trained interviewers. Eighty-five per cent of the male interviews, and approximately 80% of the female, were collected by only two interviewers. Only six interviewers were used for the male volume, and only four for the female study. Each of these interviewers was thoroughly trained and in continuing contact with the others. When their years of experience in handling interviews with every sort of individual are considered, one can assume that procedures were more uniform than might appear, in spite of the large measure of freedom to exercise judgment from case to case.

Kinsey presents a variety of evidence bearing on the quality of the interview data. One hundred and twenty-four females were reinterviewed after a minimum time lapse of 18 months. Compared with other studies of the reliability of interview reports over short spans of time and on issues much less emotional than sexual behavior, the extent of agreement between the first and second interviews is remarkably high (for a summary of such studies, see **2**). Validity of response is examined by comparing the results obtained from separate interviews with 706 pairs of spouses; here, by definition, any disagreement in reports on marital history and behavior would constitute invalidity. Again, the general findings are good and, compared with validity studies in the past interview literature, they are often amazingly good, (see **3** for a general summary). It should be noted, however, that such validity checks have been confined to spouses whose joint sexual behavior is more or less "sanctioned." Whether or not the same degree of validity holds for less sanctioned areas of behavior, such as the homosexual or extramarital, is not demonstrated.

One consideration in evaluating the findings was the fact that all of the interviewers were males. There is considerable evidence for many types of studies that the group membership (e.g., color, class) of the interviewer may seriously affect the responses. And men and women have been observed to give different replies, on many questions, depending upon the sex of the interviewer (cf. **2**). True, such studies have also shown that the more experienced and capable interviewers have been able to some extent to overcome the effects, and we would expect such effects to be at a minimum with interviewers of the caliber of Kinsey and his associates. But it has been established that statistical results can vary by as much as 10 percentage points, depending upon the sex of the interviewer, and one has no way of measuring the possible effects in this study.

The quality of the reports. A variety of criticisms of Kinsey's research have been advanced, and will continue to

be advanced, on sheer axiomatic grounds. For example, the claim is made that data concerning the most intimate of human experiences, a portion of which may be socially disapproved or remote in time, cannot, when collected by straightforward question-and-answer procedures in a relatively short interview, possibly be accurate.

With respect to criticisms based on axiomatic grounds, however, it is proper that the burden of proof rest on the critics. Kinsey himself cites considerable empirical support for the quality of his material. Reliability is high, as measured by re-interviews after a considerable lapse of time. Agreement in the independent reports of spouses concerning their joint behavior is high. There is some evidence that even events which occurred many years ago and are remote in the respondent's memory are nevertheless enumerated with considerable accuracy. Thus, Kinsey's findings on age of menarche, and on other indices of physical development, as reported retrospectively by his respondents, agree closely with independent studies based on direct observation of these matters. High agreement is demonstrated between many of his findings and the findings of earlier independent investigators who relied on different techniques. The material from the reinterviews and from the interviews with spouses concerning matters remote in time is somewhat less reliable than their reports of more recent events, but is still surprisingly good.

Terman, in his distinguished review of the earlier volume, noted that Kinsey had not taken proper account of possible memory errors in the way he *processed and analysed* his data (4). Terman did not reject

the material collected by remote recall, but argued against Kinsey's procedure of describing sexual behavior at given ages by lumping the reports of respondents actually at that age at the time of interview with the reports of those who had to recall the same period from their early lives.

In the computation of mean frequency of masturbation at age 15, for example, the memory report of a 50-year-old counts as heavily as the report of a 15-year-old . . . . It would have been helpful if he had shown in the tables what proportion of the N at that given age level was accounted for by subjects at or near that age (4, pp. 446, 450).

In the present volume, Kinsey again lumps retrospective and current reports, but he has this time provided basic tables showing the age distribution at time of report, not only for his total sample of females, but also for particular subgroups. The reader can now weigh for himself the vulnerability of certain of the data in the light of possible memory distortions.

Types of data collected. Critics have often concerned themselves with Kinsey's definition of the domain of sexual behavior and his emphasis on orgasm as the unit of measurement of such behavior. By and large, Kinsey has approached his problem in these terms. However, it seems proper to distinguish legitimate and explicit research limitations from errors which merit criticism. Kinsey frequently acknowledges that measures such as orgasm do not encompass the entire domain of the problem, but he argues persuasively that such measures were the best he could accomplish at this time. Stronger criticism, however, may be made when conclusions overreach the definition of the domain and the character of his data. Kinsey may be somewhat vulnerable on this

score. In his first volume, virtually no direct data were collected on the psychological processes underlying the reported overt male sexual behavior, and without such data some of the findings were uninterpretable. In the present volume, wherein Kinsey seems even more concerned with interpreting the underlying meaning of his findings, with commenting on psychological processes as they differentiate males and females, and with drawing recommendations for social action, there is all the more need for such auxiliary data. Kinsey has become acutely aware of this need. In contrast to the earlier report, the present study makes considerable use of attitudinal and psychological data collected directly from the interview. A few illustrations will show how the new psychological material is a basis for better inferences than the former bare record of overt behavior. Following a discussion of homosexual behavior, for example, there is a tabulation of the degree of regret, if any, that the subjects felt concerning their homosexual experience. Following a discussion of masturbation, evidence is presented on the degree to which psychological disturbance had accompanied the activity. Following a discussion of sexual contacts between adults and children, information is cited on the psychological reactions of the children.

In addition to these new data, there is acute use of auxiliary materials from an immense variety of sources in other fields. Anthropological, biological, psychological, medical, historical, and literary data on sexual behavior are profusely and effectively cited to support the findings and interpretations drawn from Kinsey's own original materials. The expanded use of such supplementary data in this current report appears to be a major contribution toward the placement of Kinsey's work in perspective.

Analytic techniques. The analytic procedures employed for the female report conform closely to the pattern established in the first volume. In general, sexual behavior is described by the incidence and frequency of *any* experience, and experience leading to orgasm. These are considered under the following defined types of sexual activity: preadolescent sex play, masturbation, nocturnal dreams, heterosexual petting, premarital coitus, marital coitus, extramarital coitus, postmarital coitus, homosexual contacts, and animal contacts. The individual figures for each type of activity, and the "total sexual outlet" from all types, are then explained by examination of the variations in such behavior according to age, religion, educational status, marital status, etc. These explanatory factors are the same ones used in the analysis of the male data, and the approach proves fruitful, for the differences reported from one group to another greatly increase our understanding of the factors affecting sexual behavior.

These statistical analyses are handled carefully. Kinsey examines his findings for spuriousness and systematically controls factors that might be correlated with the characteristic under study. When he has insufficient cases to pursue such analyses, he is careful to qualify his conclusions. Similarly, he is cautious about imputing causality.

Presentation of the detailed breakdowns by subgroups has additional values. First, it enables the reader to appraise the probable effect of the sampling biases which have been noted above. It is on this basis that

Kinsey argues that the heavy preponderance of college women is not overly serious, since, in general, he finds little difference among the high school, college, and graduate school groups. He does, however, state that "major changes might have been introduced into the generalizations if we had had a larger sample of females who had never gone beyond *grade school*" [italics mine] (p. 57).

The use of the same social characteristics to examine the sexual activity of both males and females enables the reader to compare not only the differences between men and women in the aggregate, but also differences between men and women within specific subgroups, e.g., educated, Catholic, etc. This is obviously a much more precise approach to the evaluation of what is a major social psychological contribution of Kinsey's second report—the analysis of social conflicts which may arise because of the discrepant sexual patterns of male and female partners. In marriage and in social life generally, people normally interact within limited social groupings. Consequently, the important facts to be adduced are the patterns among men and women who are likely to meet, rather than the patterns for men and women in general.

Although these parallel analyses by the same social groupings are therefore potentially valuable, Kinsey does not seem to make sufficient use of them for their obvious purpose. Most of the conclusions about social conflicts arising from discrepant sexual behavior in men and women seem to be predicated on the aggregate samples. For example, Kinsey finds that the total number of men who at one time or another engaged in premarital coitus approximates the level

of 90%, while among women the figure is 50%. He uses this discrepancy in discussing the rise of the double standard and regards the greater sexuality of young males compared with young females, as creating social strains. But when the incidence of premarital coitus is examined for separate educational levels, the figure drops to 68% among the college-educated men and rises to 60% among the similarly educated women. Within this particular group, therefore, the sex difference tends to disappear. Unfortunately, Kinsey generally tends to rest his conclusions on the characteristics of the *total* samples. It should also be noted that while it is *theoretically* possible to make detailed subgroup comparisons between the sexes, the task is extraordinarily difficult for a reader.

While parallel use of the same social characteristics in the analysis of the behavior of both sexes has very substantial value, it is strange that the female analysis should have been so rigidly restricted to those factors alone. There are other social characteristics of American women presumably central to their behavior which are not comprehended at all under the traditional groupings that were used for the male. Obvious factors in female behavior are the presence or absence of motherhood, the care of children, and status as an independent wage earner as contrasted with the role of dependent nonworking wife or daughter. Conceivably, these factors have no relation to sexual behavior, but the presumption that they do is strong. It is surprising that such analyses are not reported in the present volume.

At the outset, it was emphasized that it is not easy to make a simple evaluation of Kinsey's new work. The report represents, in this re-

viewer's judgment, a considerable improvement over the first volume. Like all investigations, it has limitations and weaknesses but, as a pioneer investigation of such scope into a very difficult research problem, it represents an unusual achievement. Certainly, one cannot accept the findings as the ultimate, definitive answer to problems of sexual behavior, but the discrepancy from this research ideal is small. Kinsey's new volume should be regarded as a monumental contribution for which all psychologists can be grateful.

REFERENCES

1. COCHRAN, W. G., MOSTELLER, F., & TUKEY, J. W. Statistical problems of the Kinsey report. *J. Amer. statist. Ass.*, 1953, 48, 674.
2. HYMAN, H., COBB, W. C., FELDMAN, J. J., HART, C. W., & STEMBER, H. *Isolation, measurement and control of interviewer effect.* Chicago: Univer. of Chicago Press, 1954, in press.
3. PARRY, H. J., & CROSSLEY, H. Validity of responses to survey questions. *Publ Opin. Quart.*, 1950, 14, 61–80.
4. TERMAN, L. M. Kinsey's "Sexual behavior in the human male": some comments and criticisms. *Psychol. Bull.*, 1948, 45, 443–459.

PART II

THE BIOLOGICAL DATA

Kinsey and his associates present in the third and final part of this volume a comparison of the sexual activity of the male and female " ... to discover some of the basic factors which account for the similarities and the differences between the two sexes" (p. 567).

The comparison is not restricted to the data reported in the first volume and in the earlier sections of the present one. Instead, it also includes supplementary data and a survey of the literature. The survey is selective and interpretive, and it represents the authors' conception of the nature of sex differences rather than a mere identification of what is known in this field. By their laborious and extensive study of one of the most important aspects of human activity, Kinsey, Pomeroy, Martin, and Gebhard have earned the interest of a wide audience in what they have to say about sexual life. Their statements need not be supported by any data, provided that their views are represented as theory, opinion, or conception. However, insofar as they claim to have demonstrated the validity of their conceptions, the reader is entitled to question whether or not this is indeed so.

Their writing is skillful and engaging, but their style is more popular than scientific. They overgeneralize from their data. They take cognizance of opposing viewpoints or contradicting data, and then proceed to advance their own viewpoint in complete isolation. A surprising number of important terms remain undefined, e.g., psychological factors, basic sex differences, sexual capacity, and capacity to be conditioned, among others. A lack of conceptual clarity is the consequence. It is necessary to interpret what they mean, since the meaning is often not evident. A good example is the summary statement in their chapter on anatomic factors in sexual response and orgasm.

In brief, we conclude that the anatomic structures which are most essential to sexual

response[6] and orgasm are nearly identical in the human female and male. ... If females and males differ sexually in any basic way, those differences must originate in some other aspect of the biology or psychology of the two sexes. They do not originate in any of the anatomic structures which have been considered here (p. 593).

The precise meaning of this statement is that the erotic topography of male and female genitalia will not account for the higher reported frequency of orgasm of male volunteers.

The authors have singled out difference in orgasm frequency as the "basic" sex difference. They could have considered other differences as "basic," e.g., gross anatomic differences, differences in degree of aggression, or attitudes toward children. They did not, perhaps because they believed the most valid data they could obtain and interpret were reports about orgasms. Initially, orgasm was represented as the prime criterion of sexual life, but subsequently became identified with all of it.

There are still other difficulties with the authors' account of the basic similarities and differences between the two sexes, and these will be discussed within the context of five main propositions. They propose that: (*a*) sexual responses depend upon a "basic" anatomy which is essentially the same in the female and the male; (*b*) the physiological accompaniments of orgasm are (with minor exceptions) basically the same for the two sexes; (*c*) males are more readily aroused by sexual stimuli because they have been conditioned more frequently than the female; (*d*) males are more frequently conditioned because certain unidentified

[6] They define sexual response as those physiological changes which lead to orgasm (p. 594).

structural characteristics of the cerebral cortex give them a greater capacity to be sexually conditioned than the female; (*e*) there are sex differences in changes of frequency of orgasm with age, the levels of the 17-ketosteroids correlating with these differences.

They dismiss the contribution of anatomic factors to an understanding of sex differences in orgasm frequency on the basis of the following arguments: (*a*) the clitoris is the embryological homologue of the penis and is erotically as sensitive; (*b*) the vagina is relatively insensitive to touch and is therefore erotically unimportant; (*c*) female homosexuals prefer clitoral stimulation, and they should know what is most stimulating; (*d*) the preferred female masturbatory technique involves clitoral rather than vaginal stimulation; (*e*) there is no evidence for sex differences in the distribution of end organs of touch and sensory nerves.

The argument of the embryologic homologue is hardly pertinent. The fact that genital differences are difficult to identify in the embryonic stage is no guarantee that *functional* differences will not emerge at a later stage. The bills of the shrike and the hummingbird are homologous structures, but function in quite different ways.

Their denial of erotic importance to the vagina is based, in part, on a study in which they asked five gynecologists to test the tactile sensitivity of the clitoris and other parts of the genitalia of nearly nine hundred women. The gynecologists used a glass rod applied lightly to test touch sensitivity. According to this study, the vagina appeared relatively insensitive to touch.

The pertinence of such a study hinges on whether the equation of

tactile and erotic is acceptable. This equation is questionable for three reasons:

1. Tactile stimulation is always available in clothed males and females without chronic erotic arousal.

2. Even tactile sensitivity may change when genital tissue is engorged by sexual excitement. There is no indication that the gynecological exploration was accompanied by sexual arousal.

3. Subjectively, tactile and erotic sensations are different.

The same data are also used to contradict the view advanced by Freud and others that the processes of maturing psychosexually involve a shift of the location of the dominant erotogenic zone from the clitoris to the vagina. This shift the authors believe to be biologically impossible (p. 584). The submitted evidence has questionable relevance since what is required is an age comparison to determine whether the clitoris remains the dominant erotogenic zone. No age comparison is provided, but rather they cite as additional support the preference for clitoral instead of vaginal stimulation by women engaged in homosexual and masturbatory activities. This type of testimony can only complicate rather than clarify a problem in which the critical independent variable is psychosexual maturity.

Their whole discussion of anatomical differences in the distribution of end organs of touch, of the density of sensory nerves in the penis and clitoris, in the size of the breast, and of other parts of the body has no bearing on an understanding of the sex difference in orgasm count unless these anatomic factors can be shown to affect orgasm-*seeking* behavior differentially. This assumption is implied but not examined.

In view of the authors' belief that the vagina is "of minimum importance in contributing to the erotic responses of the female" (p. 592), and since coitus necessarily involves vaginal stimulation, it is curious that they did not conclude that anatomic factors contribute to sex differences in orgasm count.

In discussing the physiology of sexual response and orgasm, they review certain muscular, vascular, and glandular responses before, during, and after orgasm. They base their review on a survey of the literature, references to some of their own interview data, and what is apparently a substantial amount of undocumented personal observation. While cited studies show rises in pulse rate, blood pressure, peripheral blood flow, forced respiration, salivary secretions, general muscular activity, etc. during orgasm, the evidence bearing on sex differences in these responses at the human level is variable and negligible. The authors conclude that female and male are quite alike as far as the data yet show in regard to all of these changes. While the cited physiological accompaniments of orgasm have intrinsic interest, their relevance to orgasm-*seeking* behavior is again assumed and not critically examined.

The sex difference in speed of orgasm receives special consideration as an attribute of physiological response. They state:

There is a longstanding and widespread opinion that the female is slower than the male in her sexual responses and needs more extended stimulation in order to reach orgasm. . . .

Certain it is that many males reach orgasm before their wives do in their marital coitus, and many females experience orgasm in only a portion of their coitus . . . but our analyses now make it appear that this opinion is based on a misinterpretation of the facts (p. 625).

After comparing the speed of mas-

turbation of males and females, they state that " . . . the female is not appreciably slower than the male in her *capacity* [italics mine] to reach orgasm" (p. 626).

Then they conclude: "But because females are less often stimulated by psychological factors, they may not respond as quickly or as continuously as males in socio-sexual relationships" (p. 641).

It is clear from the above that Kinsey and his collaborators are not disputing the existence of differences in speed of sexual response in a heterosexual relationship. Rather, they deny a difference in "capacity" to reach orgasm. Apparently, capacity is what is left after the influence of experience is separated out. This view emerges in the following quotation:

> For instance, the exceedingly rapid responses of certain females who are able to reach orgasm within a matter of seconds from the time they are first stimulated, and the remarkable ability of some females to reach orgasm repeatedly within a short period of time are *capacities* [italics mine] which most other individuals could not conceivably acquire through training, childhood experience or any sort of psychiatric therapy. Similarly, it seems reasonable to believe that at least some of the females who are slower in their responses are not equipped anatomically or physiologically in the same way as those who respond more rapidly (p. 377).

High-speed, high-frequency orgasm apparently represents a constellation of organic qualities which, like germ plasm, are relatively immune to the vicissitudes of experience. However, high-speed, high-frequency orgasm may be given a variety of psychological interpretations including denial of homosexual feelings, feelings of inadequacy, and expressions of guilt or exhibitionism, among others. Accordingly, the reviewer believes it is unwarranted for the authors to dismiss sex differences in speed of

response by first separating psychological factors from the performance, and then using speed of masturbation as the metric for sexual "capacity."

Having eliminated anatomical and physiological factors (exclusive of hormones) as an explanation of sex differences in orgasm frequency, the authors then ascribe the main role to psychological factors. Conditioning carries the heuristic burden. Men are more often conditioned by their sexual experience (p. 649). This view is deduced from a sexual comparison of reports of arousal in thirty-three situations. These situations include observing the opposite sex, one's own sex, portrayals of nude figures, own genitalia, etc. In 29 out of 33 of these situations, men reported a higher incidence of arousal than women. It is difficult to assess how many of these reported differences reflect a greater female sensitivity to our restrictive codes on such matters, a sensitivity which may have survived the two-hour interview with the Kinsey staff. For example, consider the four out of thirty-three sets of data in which the sex differences either disappeared or were inverted. A higher incidence of women than men reported arousal from observing moving pictures. The incidence of arousal was about the same for men and women when observing their own sex, reading romantic literature, and when being bitten. These questions appear to the reviewer to be either less loaded for women, or are equally or more heavily loaded for men. For example, to report arousal from observing a nude of the same sex implies an admission of homosexual feelings which would be equally undesirable for both sexes.

Let us assume that the reports are uncontaminated by socially determined sex differences in what is ac-

ceptable sexual behavior and treat them as valid data on sex differences in arousal. How do the authors explain the difference? For them to state that the male is conditioned by sexual experience more frequently than the female is no more than a reification of their findings. We need to know why they condition more frequently. Is it due to a socially more tolerant attitude to the male when he seeks such experience? Is he more intensely interested in these experiences? Does his greater aggression lead to a more active expression and gratification of sexual interests? The explanation of Kinsey and his associates does not become obvious until they discuss neural factors in sexual response where the following statement appears:

Since there are differences in the capacities of females and males to be conditioned by their sexual experience we might expect similar differences in the capacities of females and males to be conditioned by other, non-sexual types of experience. On this point, however, we do not yet have information (p. 712).

Apparently, from the cited statement, the authors believe that the higher frequency of human male sexual conditioning is due to his greater "capacity" to establish an association between substitute and reliable sexual stimuli, and that this capacity might extend to other types of reliable stimuli as well.[5] This explanation (stated as a finding) seems farfetched. It assumes that the male has a greater capacity for associating a class of substitute stimuli which are not peculiar to sex (since almost any stimulus can be associated with sexual response). There is no evidence for an intrinsic superiority of the male in associating in any field.

There are more plausible explana-

[5] I am assuming that conditioning is a form of association.

tions. The social code for the human female may be more restrictive, she may fear pregnancy, she may not *feel* the need for intercourse as often as the male, the male may be more aggressive—to mention a few. Kinsey and his associates mention these possibilities in one connection or another, but they reject them in favor of their "capacity" theory.

In discussing the locus of the neural mechanisms of sexual response, the authors point out that components of the response pattern, tumescence and ejaculation, are possible without sensation in paraplegics. The organism can thus function at the reflex level via the spinal cord without cerebral intervention. But while the sympathetic and parasympathetic nervous systems are also involved in the normal situation, there is no definitive evidence concerning the role of the hypothalamus.

The evidence on the role of the cortex is contradictory. On the one hand, a study on prefrontal lobotomy patients 3.3 years before, and another 3.7 years after psychosurgery, shows no appreciable deviation in orgasm frequency from corresponding normal age and sex groups. On the other hand, the authors refer to data which show that damage to the cerebrum as a whole, and particularly to the cortex, may reduce an animal's capacity to react to psychosexual stimuli. The reduction is directly proportional to the extent of damage to the cortex. In spite of the contradictory evidence, they conclude that "although the data on the relation of the cortex to sexual behavior are limited, they do show that this is the part of the nervous system through which psychosexual stimuli are mediated" (p. 712).

There are few performances indeed which cannot be credited, with

equal justification, to the cerebral cortex. The authors find support for their position from the demonstration that the male rat's copulatory behavior (and that of other animals) is more dependent on an intact cortex than that of the female. His performance requires more direction, coordination, and initiative. These differences make the male's copulatory behavior more vulnerable to cortical damage, but it does not follow, even in the rat, that cortical differences are the cause of their sex differences.

The final chapter is more carefully written. In it the authors address themselves to the problems of sex differences in the effects of aging on frequency of orgasm, and whether or not these differences have hormone correlates. Table 1 condenses and reflects the data they provide on age trends.

TABLE 1

WEEKLY FREQUENCY OF ORGASM FROM ALL CAUSES

	Age 15	Age 50
Single male	2.3	1.1
Single female	.5	.4
	Age 20	Age 50
Married male	3.2	1.3
Married female	2.2	.8

What hormone changes correlate with these differences? Data on hormone assays were not obtained from Kinsey respondents, but from studies reported in the literature. Estrogen and androgen levels do not correlate with the Kinsey data. The changes of 17-ketosteroid levels with age show a closer relationship, but there are marked deviations in the age range from 15 to 27, making a simple causal relationship untenable.

In addition to the hormones mentioned, they consider the effects of hormones from the pituitary, the thyroid, and others from the adrenals. While the latter may affect sexual behavior, it is largely as a result of their influence on the general metabolic level.

They conclude that no hormone affects sexual preferences, interests, or techniques in a selective way, but may rather modify the general level of sexual activity.

To sum up, the conception of the nature of sex differences which Kinsey and his associates have elaborated is at best awkwardly stated and misleading in many particulars. It would be wise to distinguish sharply between the valuable data that they collected and their interpretation of sex differences. While the authors acknowledge repeatedly the importance of psychological factors in understanding sex differences in orgasm frequency, the constructs that they have made central to their presentation are sexual capacity and conditioning capacity. Since they think of capacity as something devoid of experience, we are left with the question, "What do they mean by psychological?"

JONES, ERNEST. *The life and work of Sigmund Freud*. Vol. I. *The formative years and the great discoveries: 1856–1900.* New York: Basic Books, 1953. Pp. xv + 428. $6.75.

This volume is a thorough, careful, and erudite account of the origin of psychoanalysis, its invention and design by the greatest of psychoanalysts, and written by the only member of his privy council still living. The choice of this biographer was not only fortunate but almost essential, for he is the one remaining member of those six whom Freud chose after the first world war to keep alive his radical new orthodoxy. Twice Freud had destroyed his letters, diaries, manuscripts, and notes in order to preserve his personal privacy, once with a chuckle as to how much work he was making for his biographer (p. xii). But others had kept letters; the elaborate correspondence with Fliess (1887–1902) was available, and the Freud family, at first disposed to protect Freud's privacy, finally put everything at Jones's discretion, because they thought that truth was to be preferred to the false legends that were growing up about this controversial figure. Jones was thorough, as 263 footnotes and 664 citations of sources attest.

The volume describes, of course, only the immature Freud, the intense, emotional, insecure, frustrated, ambitious, impractical Freud, the Freud obsessed with making his place in his profession and in life and, as it proved ultimately, in history, the impoverished Freud too poor to marry his fiancée and for a time subordinating other ambitions to achieve

that end. This was the Freud of 1880–1900. Six-sevenths of the volume treats of these two decades. Two other volumes are promised for the four decades that remain. So often Jones remarks on the contrast between this hard-driving, insecure, uncertain man and the wise, assured, tolerant, understanding savant who was the only Freud his disciples ever knew. This first book, however, does not show how Freud wrought his own maturity by self-analysis from 1897 on. It discusses the analysis and lets the results remain to be seen in later volumes, leaving the life far short of culmination and this review merely a comment on an initial achievement whose meaning and value in 1900 were to be assessed by events that were yet to come.

It is not easy to perceive the whole Freud for any one year within these two intense decades. The biographer has chosen to devote separate chapters to different themes with greatly overlapping periods. Six chapters, for instance, deal respectively with six more or less synchronous histories.

1. First there is Freud's brief medical career, following his unduly prolonged study as a medical student, the period in which he acquired ambition (1882–1885). Actually it was his intense desire to marry that awoke professional aspiration in him, but Freud's relation to his fiancée is left for a later chapter.

2. Then there is Freud's discovery of the medical properties of cocaine and its value for internal use as an analgesic and euphoric (1884–1887). Incorrectly Freud, filled with enthusiasm for his new finding, advertised cocaine as not habit form-

ing, and lived to regret this ill consequence of his impetuosity. He missed, however, making the greatest discovery, the use of cocaine externally by injection as a local anesthetic. He had handed the preliminary indications of this application over to someone else, who presently and quite properly got the credit. Freud seems to have been blocked in pushing this line of research, perhaps because he had no taste for manipulatory experimentation or perhaps because his desire to see his fiancée after a long absence won out over his staying in Vienna to prosecute these obvious experiments that would, as it turned out, have advanced his career.

3. Another chapter in the book deals with Freud's betrothal, persistently frustrated by his poverty (1882–1886). His fiancée could expect no sizable dowry, and Freud seems at the first to have been driven to professional work in order to get enough money to marry. Again and again indigence prevented him from visiting his betrothed or being able to give her presents. When you read of these years in this chapter, you could think that wishing and planning for marriage and writing love letters took up all of Freud's time, but you would be wrong. He was pursuing half a dozen other activities simultaneously.

4. The chapter on Freud's marriage (1886) shows how need reduction quickly brought satisfaction, and thereafter parenthood (six children) and a ready lifelong monogamy. One cannot help believing that Freud was strengthened in support of his own theory of sexuality and in withstanding the *odium sexuale* that was directed toward him, because his own sexual desires were essentially conventional, both as to monogamy and as to male dominance.

5. The chapter on the neurological period depicts Freud at first still convinced that histology and morphology are the best biological disciplines, the most scientific, a view due in part to the influence first of Brücke upon Freud, and later of the histologist Meynert (1883–1897). Freud was by early preference a physicalist. He preferred organic to functional explanation and disliked what was in those days called psychology. Brücke was one of the four pupils of Johannes Müller who had declared themselves vehemently against vitalism in the 1840's. Helmholtz was another, and Jones throughout this book speaks of the influence of the "Helmholtz school" as directing Freud away from the vagaries of psychology toward the hard facts of physiology and the harder ones of anatomy. Anatomy, however, did not work out then any better than now to provide an understanding of human thought and conduct, and Freud's writing of his volume on aphasia in 1891, after he had been exposed to the magic of Charcot in 1885 and had transferred his loyalty from Meynert to Charcot, witnessed his acceptance of functional explanation when organic factors could not be found. Later, of course, he turned to psychology, still recognizing the values of the Helmholtz school by characterizing the new psychological principles as "mechanisms."

6. Then there is Freud's long and intimate association, both personal and scientific, with Breuer, fourteen years his senior. Breuer believed in Freud. He lent him money that was seldom repaid. Both of them were investing Breuer's money, Freud believed, in the promising enterprise of Freud's thought and research. At first the two men were associated in work, and it was Breuer who discovered the cathartic method and

thus in a way began psychoanalysis. Later Breuer, frightened by the phenomenon of transference in a woman ~~~~~~ escaped from the field, leaving Freud, who had greater persistence and no sense of guilt when patients fell in love with him, to carry on. Much later, when Fliess had replaced Breuer as the father-image in Freud's life, Freud turned to bitter criticism of Breuer in his letters, but he never attacked him in print.

So there you are. If you pick out the year 1885, for instance, you find Freud eating his heart out to get married, yet with time and energy left for hospital work, for his cocaine research, for his study of the neurology of the medulla, for his work with Breuer after the cathartic method had become available. Any other year is equally full. Jones, it seems to me, fails to make you feel the intensity and range of Freud's avidity; yet you can figure it out for yourself if you synthesize the chapters.

The immature Freud needed a father-image, someone to whom he could defer and whose approval and support he could win. He had three in succession during his professional life: Brücke, Breuer, and Fliess. After that his self-analysis rendered him strong enough to get along with disciples only.

It was from Brücke, rigorous physiologist and neurologist, that Freud got the faith in neurological determinism which later in the *Traumdeutung* of 1900 he turned into psychic determinism.

Breuer was the faithful admiring friend and collaborator. He gave Freud much and required little. He could not, however, bring himself to continue collaboration, and later, as Freud's theory of sexuality began to dominate his professional thought, a coolness developed between the two.

After Breuer, Fliess! Nowadays we should call Fliess a brilliant numerological crank. He believed passionately that all living matter, even the individual cell, is bisexual, that living nature is based on periodicity, and that the basic periods are 28 (the female period, the menstrual cycle) and 23 (the male period, the interval between menstrual cycles). Freud, no mathematician, was captivated by Fliess's ability to use mathematics to compute conclusions out of these basic numbers, but actually what the two gave each other was mutual admiration and the self-confidence that goes with appreciation and the appearance of understanding. They wrote to each other at least once a week, criticized each other's manuscripts, got together in important special meetings (they called them "congresses") whenever they could. Freud played the dependent role, and, when the break came later, Fliess initiated it. This was an intense 15-year friendship (1887–1902), which illustrates well the operation of human needs but does little to dignify the character of Freud. No wonder he destroyed his half of this correspondence after his self-analysis. Freud's to Fliess were saved.[1]

You have to do a little guessing and a little reasoning to find out how Freud worked and what it was he called *work*. He made "discoveries," and the subtitle of the book is "The Formative Years and the Great Discoveries." You know *that* he worked; that he worked hard; that he worked best when he did not feel well; that the intensity of his work often wore him out; that his successes were capricious, for there were long times

[1] And have been published since this review was written. FREUD, SIGMUND. *The origins of psychoanalysis: Sigmund Freud's letters to Wilhelm Fliess* New York: Basic Books, 1954. $6.75.

when no insight would come and then suddenly a particular problem would find itself resolved. He seems irregularly to have had these insights, then to have formulated hypotheses, and then to have adjusted and reshaped them as he kept applying them to his patients in analysis and in former analyses. He seems almost never to have accepted a new hypothesis quickly, but only slowly and, as it were, reluctantly, after much checking. Nor did his accepted hypotheses always stay right. There was Freud's belief about adult hysterical symptoms' always having their roots in infantile sexual experience. What at first seemed to confirm this belief was the constant report in analyses of incestuous relations between parents and children. Freud, as well as Breuer, had difficulty in accepting this finding as a fact, and then suddenly in 1897 Freud realized that these attempted seductions were nothing more than patients' fantasies that had never really occurred. That "discovery" was a blow to Freud, for it seemed to knock out half of the structure of psychoanalysis, leaving intact only the theory of dreams as wish-fulfillment. Actually, as it turned out, the "discovery" strengthened the theory of infantile sexuality, weakening only the false belief in parental sexuality.

Although Freud usually was slow to accept these insights, once in a while an insight was sudden enough to be dated. One such instance is indicated by Freud's suggestion, in a moment of gaiety, that there should be erected a marble tablet at the table near the northeast corner of the terrace of Bellevue Restaurant in Vienna. The tablet was to read: "Here the secret of dreams was revealed to Dr. Sigm. Freud on July 24, 1895" (p. 354).

Freud's technic was not, of course, the experimental method. He himself remarked in 1900: "I am not really a man of science, not an observer, not an experimenter, and not a thinker. I am nothing but by temperament a *conquistador*—an adventurer" (p. 348). And at another time: "It seems to be my fate to discover only the obvious: that children have sexual feelings, which every nursemaid knows; and that night dreams are just as much wish-fulfillment as day dreams" (p. 350). On the other hand, one should not take Freud's modesty and momentary self-depreciation too literally. He thought of himself as a discoverer, and those insights that stayed valid were the result of long and laborious thinking within the monotony of psychoanalyses and their repeated review. Freud's technic lay somewhere between that of the experimental psychologist, who alters an independent variable and observes the result, and the philosopher-psychologist, who induces generalizations about human nature from the reservoir of his experience. Freud made his generalizations from a wealth of specialized experience and then tested his hypotheses out against particular cases, increasing his assurance about the validity of each induction as the number of consistent cases grew. He had, however, no control, either in the sense of the rigorous constraint of contributing factors or in the sense of adding the method of difference to the method of agreement. Indeed he seems to have been restricted to Mill's method of agreement, pure and simple, a method which by itself is clearly unsafe.

Was Freud original? He fitted his *Zeitgeist*, of course, as all great men do. Brücke influenced him in respect of determinism, and Meynert in

respect of active ideas and their repression, for Meynert was an admirer of Herbart, who in these matters is clearly an intellectual ancestor of Freud's. Jones mentions as also influential Fechner and Griesinger, both of whom derived many of their ideas from Herbart. Breuer's influence we have already noted. Charcot captured Freud's admiration when he visited him in 1885 and later Bernheim became a lesser influence. Charcot it was who led Freud to shift his allegiance from the organic to the functional and thus away from Meynert. Jones does not find that Brentano influenced Freud, although Freud knew him, sat under him in lectures, translated a volume of J. S. Mill at his recommendation. There is no relation between Janet and Freud except that both were affected by the same *Zeitgeist* and by Charcot.

Yet all this fitting of Freud into his intellectual genetics is not very important. Certainly Freud was vastly original. His conceptions of infantile sexuality and of the psychological mechanisms are not to be found in Charcot or elsewhere, and the importance of his originality is to be measured by its effects. In this reviewer's judgment Freud is a great man because he influenced thought and civilization more than any other person whom psychology might claim in the last three hundred years. A Tolstoy might have shown him as more the agent of an irresistible *Zeitgeist* than as a great originator, and still he would be a Great Man, for that is what Great Men are. Kings are History's slaves. That the kings who have commanded the choices progress should make are singled out by posterity to be called great is the proof of man's essential insecurity and his need to find safety in pride and reverence for others than

himself. Freud would have recognized such a fact without losing his faith in Helmholtzian determinism; he would have believed that greatness is not diminished by its having causes.

<div align="right">EDWIN G. BORING.</div>
Harvard University.

FLOYD, W. F., & WELFORD, A. T. (Eds.) *Symposium on fatigue.* London: H. K. Lewis & Co., 1953. Pp. vii+196. 24s.

This report of both English and American workers is concerned with various aspects of inadequate performance of the human, and techniques for detecting and measuring them. Both old and new slants are represented. While the symposium is labeled as one on fatigue, no verbalized convention is agreed upon as to what is fatigue and what is to be otherwise defined. In some cases, the participants admit that they are not really reporting on fatigue but rather upon decreasing performance (R. C. Browne). Some tacitly imply that fatigue is performance decrement (R. M. Gagné), some explicitly state it (W. T. Singleton). A. T. Welford explicitly recognizes that there are several distinctly different phenomena called fatigue. One is a subjective state related to some kind of physical or mental strain. The other is a decrement in performance following more or less prolonged activity. He states that the psychologist takes a position "between the two." This is to say that the psychologist must "keep in touch with the man in the street"; he must "build his theories in such a way that they can contain the experiences and behavior of ordinary people." Welford further states that "it is to physiology that he must look for explanations."

Others try to connect fatigue with learning and drives as defined in prevalent learning theory (D. R. Davis). It seems as though learning deserves psychological theorizing but that fatigue does not. T. A. Ryan deals very ably with the factor of *effort* in performance and thus comes closest to saying that fatigue is a proper study for psychologists along side other phenomena such as anxiety. Though he did not, Ryan might have said that since he found effort definable only in experiential terms, the *sine qua non* of fatigue is an identifiable experience, and at the same time a respectable object of scientific inquiry.

Nowhere throughout the printed report of the symposium does there seem to be a recognition that the topic of fatigue is in an undefined and therefore unusable state for intelligent discussion and inquiry. There was no recognition of the logically intolerable condition of having a number of terms used synonymously in technical discourse; e.g., fatigue, boredom, tiredness, work decrement, disorganization of performance, and impairment. Because of this deficiency and what follows from it, all these very careful and ingenious studies lack the orientation and meaningfulness that they would otherwise have.

It is to be hoped that the need to systematize knowledge and information, and to orient ourselves by systematic thinking, will soon be recognized to be as urgently required as the continued collection of more data. To close with a positive note, it can be said that the reader will find in the symposium a number of ingenious experiments that are well worth his acquaintance.

S. HOWARD BARTLEY.
Michigan State College.

HOVLAND, CARL I., JANIS, IRVING L., AND KELLEY, HAROLD H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953. Pp. xii+315. $4.50.

This volume brings together the findings of the Yale Communication Research Program, a program devoted to the study of the modification of attitudes and opinions through communication by means of controlled experiments. The project was set up as a cooperative research and study group rather than in a centralized, hierarchically organized form. Approximately 30 individuals contributed to the work, each being encouraged to pursue his own interests and theoretical bent.

Ideal as this plan may be from many points of view, it clearly has made the job of reporting a difficult one. The researches are uneven in quality, and stem from a variety of theoretical approaches, including, among others, Hull's learning theory, some motivational hypotheses of Freud and other psychoanalysts, and some of the formulations of Lewin, Sherif, Newcomb, and others concerning the effects of group membership. The authors of *Communication and Persuasion*, therefore, adopted a structural organization, using topical headings of (*a*) the communicator, (*b*) content of the communication, (*c*) audience predispositions, including group conformity motives and individual personality factors, and (*d*) responses, including overt expression of opinion and retention of opinion change. In the light of the formidable organizational task they faced, the authors are to be congratulated on the book's structural unity, ease of transition, and readability.

The experiments themselves deal almost exclusively with the effects of

one-way communication on "captive" audiences, opinion change and retention being assessed by means of questionnaires. The resulting limitations are obvious, but since most of the questionnaires were carefully worked out with numerous cross checks, and since all the experiments are primarily directed toward basic theoretical rather than immediately practical aims, it seems likely that many of the generalizations will hold for more complex communication situations evaluated in other ways. The probability of this being so is increased by the fact that the authors have made a commendably thorough review of the relevant literature, which is presented for each experiment or set of experiments as the context within which the new findings are considered. In addition, most of the experimenters have made admirable use of advanced experimental designs which allow the assessment of numerous variables at once, both singly and in interaction.

In summary, this book is an excellent report of studies that make a substantial contribution to an important area of social psychology. It is a source of many experimentally derived, thought-provoking generalizations concerning the processes of communication and opinion change. These generalizations, however, tend to be quite limited, specific, and tentative, thus reflecting the present state of knowledge in the field, the over-all eclectic nature of the research reported, and the authors' preference for sticking close to the available data. The volume also provides an excellent review and analysis of a large area of communication literature, as well as a number of valuable suggestions for future research.

F. P. Kilpatrick
Princeton University.

Bingham, Walter VanDyke. *Homo sapiens auduboniensis*. New York: National Audubon Society, 1953. Pp. 39. $1.10.

A significant, if not the most conspicuous, phase of the late Walter Bingham's qualities receives a sidelight from the address "Homo Sapiens Auduboniensis," which he gave in October 1937 at a dinner of the Audubon Camp held at the American Museum of Natural History. The address gives its title to a recent memorial volume of which it is the nucleus.

Following a brief, touching foreword in the name of the Audubon Society, there is a memoir by Walter Bingham's wife, Millicent Todd Bingham, which embodies the major portion of the 39-page text. Its title, "Beyond Psychology," is not invidious to our discipline, meaning rather, "What should they know of Bingham who only Bingham knew" for his professional accomplishment, having but fragmentary perception of his various reachings into the humanities and the arts, particularly music; not to say the depth and breadth of good will he was so apt in implementing, his steady cheeriness, his hard practical sense, his effectiveness in physical emergency.

Then come the five pages of the "Homo Sapiens Auduboniensis" itself. The pages that one of our group would least want to miss are the two at the volume's close, a (1948) letter of Bingham's to the able and energetic director of the Audubon Camp, Carl Buchheister. It concerns the lag of social behind material culture and something to do about it; compare also the note on an earlier letter to Clark Wissler (pp. 26–27). The address itself, as implied in its title, is mainly a whimsical characteriza-

tion of Audubon Camp participants, phrased in natural history terms. For Bingham's professional colleagues, this theme has perhaps less intrinsic interest than other portions of the volume, though towards the close he emerges from the role-taking more fully. The remarks have a good deal more interest for the psychographer, for the part that humor played in Bingham's personality, a sort of testing the limits of facetiousness that he would allow himself.

What was the role in Bingham's personality of the "humor" that Mrs. Bingham simply mentions (p. 15) after rather discounting it on the preceding page? In humor's ethically highest reach,[1] Bingham was at home with the best minds. In the other levels, more connoted by "sense of humor," he was progressively less at home. Fun-loving would hardly describe him, though he was ready enough that others should be so. In the stereotype of the *salon* he would have passed for a relatively humorless man. Still less was he one to deal in the Joke Proper; and Flippancy tended to disgust him. At any level, the humor of "Wit" was not Bingham's cup of tea; that of Joy emphatically was.

[1] " . . . Joy, Fun, the Joke Proper, and Flippancy. You will see the first among friends and lovers reunited on the eve of a holiday. . . . Something like it is expressed in much of that detestable art which the humans call Music. . . . " (C. S. Lewis, *The Screwtape Letters*, p. 57).

The ego-defense dynamic of wit seems well established, though it matters much what it is a defense against; compare Abraham Lincoln with Jonathan Swift. In a character like Bingham there is little need for such defense, and it does not develop readily or far. To a whimsicality like "Homo Sapiens Auduboniensis" he could bring himself; satire would have been mostly beyond him, or beneath him, as you please. Try to imagine Walter Bingham composing a *Lilliput*, let alone a *Modest Proposal*.

Bingham's professional colleagues may well thank the National Audubon Society for their graceful tribute, and for the light it throws on this comparatively unconsidered facet of Bingham's rich personality. Beyond this, of course, looms the whole question of the role of nature study in the benign socialization of the individual. It *can* be a negative one. How much did love of nature help to make Bingham what he was? How much did Bingham, being what he was, have to respond to nature in this way? We are not likely to know. What we do know is that this love of nature had no small part in one of the best and wisest men that ever made psychology his profession; that he had a lively and steadfast faith in its benign influence on others, and in its communicability.

FRED L. WELLS.
Newton Highlands, Massachusetts.

ABRAMSON, HAROLD A. (Ed.) *Problems of consciousness.* Transactions of the Fourth Conference, Josiah Macy, Jr. Foundation. New York: Josiah Macy Jr. Foundation, 1954. Pp. 177. $3.25.

ARNOLD, MAGDA B., & GASSON, JOHN A. *The human person; an approach to an integral theory of personality.* New York: Ronald, 1954. Pp. x+593. $6.00.

BAKER, JAMES, Jr. *The exteriorization of the mental body.* New York: William-Frederick Press, 1954. Pp. 32. $1.50.

BARRON, MILTON L. *The juvenile in delinquent society.* New York: Knopf, 1954. Pp. xix+347. $5.00.

BETTELHEIM, BRUNO. *Symbolic wounds; puberty rites and the envious male.* Glencoe, Ill.: The Free Press, 1954. Pp. 286. $4.75.

BLAIR, G. M., JONES, R. S., & SIMPSON, R. H. *Educational psychology.* New York: Macmillan, 1954. Pp. xvii+601. $4.75.

BURINGTON, RICHARD S., & MAY, DONALD C., JR. *Handbook of probability and statistics with tables.* Sandusky: Handbook Publishers, 1953. Pp. ix+332. $4.50.

COLE, LUELLA. *Psychology of adolescence.* (4th Ed.) New York: Rinehart, 1954. Pp. xvi+712. $6.00.

COMMINS, W. D., & FAGIN, BARRY. *Principles of educational psychology.* (2nd Ed.) New York: Ronald, 1954. Pp. xvi+795. $5.75.

DE SADE, LE MARQUIS. *Selected writings of De Sade.* (Trans. by Leonard de Saint-Yves.) London: Peter Owen, 1953. Pp. 306. $6.75.

EDUCATIONAL RECORDS BUREAU. *1953 fall testing program in independent schools and supplementary studies.* Bulletin, No. 62. New York: Educational Records Bureau, 1954. Pp. xiii+84.

ENGLISH, HORACE B. *The historical roots of learning theory.* Garden City: Doubleday, 1954. Pp. 21. $.65.

FRANK, LAWRENCE K. *Feelings and emotions.* Garden City: Doubleday, 1954. Pp. vi+38. $.85.

FRASER, D. C. *A psychological glossary.* Cambridge: W. Heffer & Sons, 1954. Pp. vi+40. 3s. 6d.

GANNON, TIMOTHY J. *Psychology; the unity of human behavior.* New York: Ginn, 1954. Pp. xii+482. $4.75.

GILLIN, JOHN. (Ed.) *For a science of social man.* New York: Macmillan, 1954. Pp. vii+289. $4.00.

GLADWIN, THOMAS, & SARASON, SEYMOUR B. *Truk: man in paradise.* New York: Wenner-Gren Foundation, 1954. Pp. 655. $6.50.

GORER, GEOFFREY. *The life and ideas of the Marquis de Sade.* (Rev. Ed.) London: Peter Owen, 1953. Pp. 244. $3.50.

HARMS, ERNEST. *Essentials of abnormal child psychology.* New York: Julian Press, 1953. Pp. xiii+265.

HONIGMANN, JOHN J. *Culture and personality.* New York: Harper, 1954. Pp. x+499. $5.00.

IKIN, A. GRAHAM. *Life, faith and prayer.* New York: Oxford Univer. Press, 1954. Pp. 127. $2.50.

JACOB, PEYTON. *The behavior cycle; an interpretation of behavior from the standpoint of an educationalist.* Ann Arbor: Joseph S. Jacob. 1954. Pp. vii+126. $2.50.

LEUBA, CLARENCE. *The natural man.* Garden City: Doubleday, 1954. Pp. x+70. $.95.

LLOYD, FRANCES. *Educating the subnormal child.* New York: Philosophical Library, 1953. Pp. vii +148. $3.75.

NUTTIN, JOSEPH. *Tâche réussite et échec; théorie de la conduite humaine.* Amsterdam: Publications Universitaires de Louvain, 1953. Pp. ix+530. 330 fr.

PEPINSKY, HAROLD B., & PEPINSKY, PAULINE N. *Counseling theory and practice.* New York: Ronald, 1954. Pp. viii+307. $4.50.

RYAN, THOMAS ARTHUR, & SMITH, PATRICIA CAIN. *Principles of industrial psychology.* New York: Ronald, 1954. Pp. xiv+534. $5.50.

SAENGER, GERHART. *The social psychology of prejudices.* New York: Harper, 1953. Pp. xv+304. $4.00

SORENSON, HERBERT. *P education.* (3rd Ed.) New York: McGraw-Hill, 1954. Pp. $5.00.

STONE, CALVIN P. *Abnormal psychology glossary.* (Rev. Ed.) Stanford: Stanford Univer. Press, 1944 1954. Pp. 24.

THORPE, LOUIS P., & SCHMULLER ALLEN M. *Contemporary theories of learning.* New York: Ronald, 1954. Pp. viii+480. $5.50.

WERT, JAMES E., NEIDT, CHARLES O., & AHMANN, J. S. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954. Pp. vii +435. $5.00.

WELBERG, LEWIS R. *The technique of psychotherapy.* New York: Grune & Stratton, 1954. Pp. xiv +869. $14.75.

Psychological Bulletin

PSYCHOLOGY IN JAPAN

KOJI SATO AND C. H. GRAHAM

Kyoto University and Columbia University

This paper presents a survey of Japanese psychology, first, from a historical point of view, and second, from the point of view of describing some of its modern aspects. The occasion for organizing the paper arose in the summer of 1952 when the two authors were members of the Seminar in Experimental Psychology, one of the Kyoto Seminars in American Studies conducted under the auspices of Kyoto University, Doshisha University, the University of Illinois, and the Rockefeller Foundation. Many of the ideas expressed and information contained in the account are products of the many discussions that went on among the Seminar participants along with the formal academic activities.[1]

The present survey attempts, within its limitations, to trace, in general outline and without claim to exhaustiveness, certain historical patterns of Japanese psychology and their emergence into the modern stream of fact, theory, and practice. The account enumerates many researches that have been carried out by Japanese investigators, researches that are not usually well known in the West. It is hoped that the present description will stimulate interest in these contributions.

HISTORY OF PSYCHOLOGY IN JAPAN

Psychology until 1926

Western psychology was introduced in Japan about 10 years after the beginning of the Meiji Era (1868–1912) against an ancient background of Indian, Buddhist, and Chinese philosophy. Once the introduction was effected, translations were quickly made of books by Haven, Bain, Sully, Wundt, and Ladd. The first experimental researches of Japanese psychologists were performed in America and Germany in the last decade of the nineteenth century by such men as Motora (111), Matsumoto (87, 88, 89, 90, 91), Nakajima (116), T. Okabe (142), and Kakise (43).

Yujiro Motora (1858–1912) and Matataro Matsumoto (1865–1943) were the principal pioneers of Japanese psychology. Motora was the first professor of psychology in Tokyo University. He had a philosophic mind and tried to establish a system of psychology. Matsumoto was primarily an experimentalist. He designed the psychological laboratory of Tokyo University in 1903, after inspecting laboratories in America and Germany, and established a laboratory at Kyoto University in 1907. He became the professor of psychology at Kyoto in 1906 and succeeded Motora at Tokyo after the latter's death in 1913. He advo-

cated the study of "psychocinematics," a science of psychophysiological behavior, and became the father of applied experimental psychology in Japan. He survived Motora by 31 years. Most of the older leaders of Japanese psychology are his students.

The first 25 years of the twentieth century constituted an orientation period in the methods of psychological research. Before 1925 studies appeared under the names of such researchers as Chiba (16, 17), K. Masuda (83, 84, 85), Yh. Kubo (64, 65, 66), K. Tanaka (183, 184, 185), G. Kuroda (70, 71), Sakuma (154, 155, 156, 157), Koga (58), Chiwa (18), and Narasaki (119). In 1911, Ohtsuki (138) published his *Experimental Psychology*, a book of 1,000 pages, which owes its material to Wundt, Titchener, and other German and American psychologists. The prevalent schools of psychology in Japan at this time were Wundt's apperception psychology and American functionalism. Students of structuralism, such as Yokoyama (218, 219) and Takagi (175), came at the end of this period, as did Yatabe, who had just returned from Piéron's laboratory.

Experimental studies had started in the psychological laboratories of Tokyo and Kyoto universities near the beginning of the century. These works were first published as monographs, and it was not until the *Japanese Journal of Psychology* was founded in 1919 at Kyoto by Genji Kuroda that a ready means of publication became available. The editorial office was moved to Tokyo in 1923.

With the advent of better economic conditions after World War I the number of educational institutions increased greatly. Departments of psychology were established in Tohoku

(Sendai), Kyushu (Fukuoka), Keijo (Seoul, Korea), Taihoku (Taipei, Formosa), Nihon, Hosei, Waseda, Keio (the latter four in Tokyo), Doshisha (Kyoto), Kwansei Gakuin University (Kobe), and the Tokyo and Hiroshima Colleges of Arts and Science. In addition, many psychologists obtained positions in the Kotogakko (essentially junior colleges).

It is not surprising that the increase in Japanese educational facilities was accompanied by a quickened interest in educational psychology, an interest that was becoming strongly established by World War I. The mental test movement found a home in Japan, but many of its technical aspects were not developed along the lines that prevailed in America.

Studies of personality did not become important in number until the 1930's, but it is important to note that Watanabe wrote a book on personality (202) as early as 1912. Industrial psychology was recognized by some people as an important field of research during World War I, but, probably due to problems peculiar to Japan because of overpopulation, it did not develop (nor has it yet developed) to the same important position it holds in the United States. In clinical psychology, the psychiatric institutes of Tokyo and Kyoto universities participated in the mental test movement and devised psychodiagnostic methods, but only a few individual psychologists, e.g., Oguma (133), took an interest in problems of abnormal psychology. Kuwata (75) of Tokyo University introduced the folk psychology of Wundt, and Iritani (36) the social psychology of McDougall.

In general, it may be said that despite the existence of certain scattered interests in other areas, psychology in Japan until 1926 was largely

dominated by educational psychology and classical experimental psychology.

One major contribution of great influence that was to have an important bearing on later developments in Japanese psychology appeared in 1925. In this year Matsumoto wrote his comprehensive *Psychology of Intelligence* (90), a book that dealt not only with intelligence, but also with mental work, senescence, environmental factors, efficiency, and military psychology. In addition it included a broad commentary on the general development of psychology. It involved a survey of research done by Japanese psychologists under his (Matsumoto's) guidance as well as a consideration of related work done by western psychologists. It was a milestone in the development of Japanese applied psychology.

From 1926 through World War II

The beginning of the Showa Era (1926) marked the development of some important trends in Japanese psychology. At that time the influence of gestalt psychology was beginning to be felt; the Japanese Psychological Association was established, as was the new series of the *Japanese Journal of Psychology;* and the number of graduates in psychology showed a remarkable increase due to the expanded system of higher education that developed after World War I.

Sakuma (157) and Onoshima (146), who studied at Berlin University, were the chief expositors of gestalt psychology. Sakuma's translation of Köhler's *Gestalt Psychology* (154) is well known. It can safely be said that gestalt theory has colored the thought of most Japanese psychologists from about 1926 to World War II (and even now exerts a powerful influence). The years around 1926 also saw the introduction of the works of the Leipzig school (e.g., those of Krueger and Volkelt) by Iwai (40, 41) and others (159) of Kyoto University. It would, however, be wrong to think that these influences were unaccompanied at this time by original Japanese contributions. Chiba proposed a theory of "Eigenbewusstsein," R. Kuroda the psychology of "comprehension," and Sakuma a theory of "basic consciousness" as directly experienced. These contributions may be seen as an attempt to revise the traditional theory of consciousness in the light of some characteristics of oriental culture.

Among books that appeared near 1926, several deserve particular mention. K. Masuda's *Introduction to Experimental Psychology*, a manual for beginners, appeared in 1926. The same author's *Methodology of Psychology*, published in 1934, is an authoritative work in general psychology. Yh. Kubo compiled two volumes of his important *Handbook of Experimental Psychology* during 1926 and 1927, and Kido's *General Outline of Psychology* appeared in 1931.

Experimental Psychology

What were some of the evident trends in experimental psychology from 1926 to World War II?

Perception. In the years between 1926 and 1939, the study of perception progressed at a quickened pace, usually along lines dictated by gestalt psychology. A short enumeration of some of the studies undertaken may be used to indicate the areas of greatest interest.

The Psychological Institute of Kyushu University published a series of experimental studies on the structure of perceptual space directed by Sakuma, Yatabe, and Akishige (8, 9,

95, 158, 182, 206). Takagi (**176**) compared the structural, phenomenological, and gestalt views of perception. Obonai (**30, 121, 125, 126, 127**) continued (and has to the present time) his studies of psychophysical induction in perception, memory, and other areas. (Most of the contributions are in the area of visual perception, where Obonai has tried to formulate laws of induction as extensions of the laws of contrast and confluence.)

Matsubayashi (**86**), an ophthalmologist, performed his classical experiments on depth in 1937 and 1938. A number of investigations of the visual constancies were carried out. Akishige (**8**), Ibukiyama (**27**), and others worked on size constancy; Yt. Kubo (**67**) on shape constancy; and Ogasawara (**130**) on the constancy of phenomenal velocity. Nishi (**121**) studied the moon illusion.

Studies on the perception of movement may be represented by the investigations of Mukuno (**113**) on tactile apparent movement, S. Kubo (**63**) on visual apparent movement, Hisata (**27**) on auditory apparent movement, Fukutomi (**21**) on delta movement, and Ogawa (**131**) on the seen path of real movement. Studies on form perception were carried out by Morinaga (**102**), Hayami and Miya (**25**), Yagi (**203**), and others. Studies of the relation between time and space (the "tau effect" and its reverse) were performed by Abbe (**1, 2**) and by Suto (**172**).

At this time Y. Wada (**200**) performed experiments on the time error for auditory stimuli. Other studies in audition were not extensive, but Yuki made an important contribution when he published *A Psychology of Tone* (**221**) in 1933. In the field of gustatory sensitivity, Rikimaru (**151**) made a survey of taste blindness among Japanese, Chinese, and Formosan natives.

The area of eidetic imagery was investigated by Ohwaki and Masaki (**80**) and the former (**139**) published his book *Eidetic Imagery of Children* in 1935.

The work in perception during this period was essentially the product of a relatively small number of Japanese psychologists. These psychologists were very active and contributed a number of results which are, unfortunately, not as well known in western countries as they should be.

Human learning, animal learning, and thinking. The field of human learning was not so intensively investigated from 1926 to 1939 as was the area of perception. The studies that were done were generally performed within the framework of gestalt psychology. For example, Ushijima (**197**) studied the achievement problem (*Erfolgsproblem*) in learning, Amano (**11**) investigated "memory traces" in the tradition of the Wulf experiment, and problems of retroactive inhibition, proactive inhibition, and reproductive inhibition were taken up by Sagara (**152, 153**) and Maeda (**76**).

Researches in the field of animal behavior were sparse, but a few interesting investigations were undertaken. R. Kuroda published studies on the hearing of reptiles as early as 1923. Thereafter he extended his researches to the monkey, white rat, and tortoise. He wrote a *Psychology of Animals* (**74**) in 1936. Takagi (**178**) studied the influence of background upon the transposition of selective responses to brightness (in the varied tit) and investigated form constancy in the tomtit (**177**). The transposition of selective responses

was also studied by Ohtsuka (137) and Shirai (164). Odani (128), a psychologist and psychiatrist, studied the effects of cerebral lesions upon the visual and auditory discrimination habits of white rats. Yamanouchi, a biologist, published his *Psychology of Animals* in 1938. Hayashi (26), a student of Pavlov, pioneered in the study of conditioning.

The general problem of thinking has been of great interest to Japanese psychologists. Kido (69), Mitui (96), and Tomeoka (189) of Hosei University performed experimental observations on the process of doubt and speculation. Sato (160), Abe (5), and Amano (12) analyzed the process of comparison by introspective methods. Sato (160) has extended his interest in this general area to investigations of developmental aspects of the comparison process.

Personality

In the thirties German characterology was introduced into Japan, and some experiments (10) were performed within the framework of this viewpoint. These experiments constituted the first systematic work in the field of personality. For example, psychologists (7, 187) at Waseda University performed research in the context of Kretschmer's typology, and Uchida (191), in particular, devised a continuous addition test, essentially a modification of Kraepelin's work. Uchida's test is now in general use in Japan (190). During this period, Abe (6), Susukita (171), and others, under the guidance of Chiba, studied personality differences among Manchurian races. Masaki and Yoda (82) studied personality from the viewpoint of educational psychology and wrote *A Psychology of Character* in 1937.

Educational and Clinical Psychology

Educational psychology, one of the two oldest fields of psychological investigation in Japan, showed considerable progress between 1926 and World War II. Studies of development, learning, personality, work, and fatigue increased greatly in number and had direct influences on education.

The Institute for Child Welfare of Aiikukai and the Institute for Child Study in Nihon Women's University played particularly important roles in researches in child psychology during the thirties. Tokyo (with Tanaka, Narasaki, Terazawa, and Takemasa [120, 181, 183]) and Hiroshima College of Arts and Science (with Kubo and Koga [57, 64, 66]) were centers of educational psychology. Kyoto University, with Nogami (123) in adolescent psychology, and Iwai, Kato, Sonohara, and Moriya in child psychology, was a center of developmental psychology. At this time Yh. Kubo (66) wrote his *Child Psychology* under the influence of Charlotte Bühler, and Hatano introduced Piaget's work (24) to Japan. The main weakness of the child studies carried out in this period lay in the fact that no provisions were made for large-scale follow-up investigations. Investigations in the areas of maturity and old age were done almost solely by Tachibana (174).

The investigation of children's personalities was taken up in the period from 1920–1935. Intelligence tests and personality tests were not elaborately developed or used in Japan until after Ohtomo, a student of Judd, wrote his two-volume *Diagnostics of Education* (136) in the period 1928–1933. A revision of the Binet-Terman tests of intelligence by H. Suzuki (173), then a school in-

spector of Osaka City, has been in general use in Japan since 1930, and group intelligence tests were also adapted for Japanese use by Tanaka (185), Kirihara (53), Awaji (13), and others. The only personality tests developed during this period were a test of extroversion-introversion by Awaji and Okabe (14), and a revision of the Downey Will-Temperament test by Kirihara.

In the 1930's psychologists, child psychologists, and educationists formed groups to study clinical and abnormal child psychology. Their studies mark the beginning of clinical psychology in Japan. Several child guidance clinics were established about 1930 (in Kyoto City, Kobe City, Hyogo Prefecture, Aichi Prefecture, and Tokyo City). At the same time a few clinics were established in prisons, courts, and reformatories under the Japanese Judicial Department and in guidance centers set up by the Welfare Department. The term *clinical psychology* was never used officially during this period, and the work of the few psychologists employed in the clinics was limited to diagnosis and the evaluation of intelligence and character. The work of the early clinicians and the work of the educational psychologists overlapped, but their respective efforts differed in one respect: the former extended their methods to meet the requirement of evaluating abnormal children and criminals. In a few cases the clinicians adopted methods typical of psychoanalysis (Marui [79] and others).

Industrial Psychology

This field, always at a disadvantage in Japan because of the everpresent overpopulation and its consequence, cheap labor, did not progress as rapidly during the 1930's as some other areas of psychology. It

has been mentioned that, during World War I, attention was paid to industrial psychology and some systematic studies of the labor problems were made subsequent to 1926 by members of the Kurashiki Institute of the Science of Labor, founded in 1921 by M. Ohara. (The Institute moved to Tokyo in 1937 and has continued its work until the present time. It has included several psychologists on its staff, Kirihara and Ueno [52, 192], for example.) The Prefectural Institute of Industrial Psychology of Osaka was founded about 1925. By the middle of the thirties the Ministry of Welfare and the Government Railways employed a number of psychologists for vocational guidance and the promotion of efficiency of labor. There have been two journal publications in this field: *The Science of Labor* and *Management*. During the war human resources were in short supply, and many books on the management of labor appeared. A considerable number of psychologists entered the Army and the Navy and took part in researches on aptitude testing and personnel problems.

Social Psychology

Wundt's folk psychology and McDougall's social psychology were studied at an early date by Japanese psychologists, but research in the social area did not develop for a long time. The measurement of attitude was first undertaken by Koga (56) about 1934, and problems of race differences (133) and group psychology (46, 47, 97) interested psychologists from about 1931 to World War II. During this period, problems of national morale were seriously taken up by only a few psychologists, due, probably, to the fact that in a Japan dominated by the military caste, such questions were considered

to be political or philosophical in nature rather than psychological.

General Observations

During the period 1926 to World War II, Japanese psychologists were self-conscious and self-critical. Their science was young, and they were passing through a period of self-examination and, to some degree, of confusion. Matsumoto, president of the Japanese Psychological Association in 1933, discussed (91) what he considered to be the three weak points of Japanese psychology: (a) lack of communication among psychologists, resulting in a decreased effectiveness of research; (b) excessive use of group methods involving questionnaires, tests, etc., and the paucity of analytic experimental researches; (c) excessive speed in adopting new, and probably transient, western psychological concepts, without, in all cases, applying appropriate historical criticisms and evaluation of facts.

The years near 1935 were the high watermark of Japanese psychology before World War II. The number of psychological journals was at a peak and psychologists were very active. *The Tohoku Psychologica Folia* (Sendai), the *Japanese Journal of Experimental Psychology* (Kyoto), and the *Acta Psychologica Keijo* (Seoul) became important scientific journals. The *Japanese Journal of Educational Psychology* (Tokyo), the *Japanese Journal of Applied Psychology* (Hiroshima), and *Animal Psyche* (Tokyo), featuring articles in a somewhat popularized style, signaled the emergence and development of new and important areas. Two journals in fields related to psychology, *Studies in the Science of Labor* and *Beiträge zur Psychoanalyse* (Sendai) became influential. According to a survey by Nakamura and Nagasawa, about 100 Japanese works were abstracted in the *Psychological Abstracts* between 1933 and 1940.

The number of reports given at the fourth meeting of the Japanese Psychological Association at Sendai in 1933 was 84; at the fifth meeting at Tokyo in 1935, 126; and at the sixth meeting at Keijo, Korea, in 1937, 63. At these meetings, strictly experimental work constituted about 40 per cent of the papers. Sensation and perception were the major interests, about one in four papers being in these areas.

Psychological research in Japan was severely reduced at the time of the outbreak of the Sino-Japanese conflict in 1937 and its later development into the Pacific war in 1941. Several journals could not be continued, and at the end of the war even the *Japanese Journal of Psychology* could not be published regularly. During the long period of strife a number of psychologists worked in the Navy and the Army, but, in general, it may be said that psychological research was almost completely interrupted. The number of graduate students in psychology dwindled to essentially zero.

From the End of World War II to the Present

After World War II a great number of changes took place in Japanese life, many of them attributable to the Occupation. Among other things, the changes include a new and liberalized Constitution and a proposed renovation of the Japanese educational system. We cannot consider here all the factors behind the influx of students into universities and the influence exerted by the "new" education. It is sufficient to say that the number of graduate students in psychology increased considerably above the prewar level, and there was

he showed interest in psychology, particularly in areas that had not been well developed previously. At the same time, the Japanese Psychological Association greatly increased its membership, and its meetings were held annually. During the 1951, 1952, and 1953 meetings, the number of reports presented exceeded 300. The percentage of the reports that could be termed experimental was about 30.

A further discussion of trends may here be aided by a subject-by-subject consideration of work in the various areas.

General Psychology

Yatabe wrote his *Introduction to Psychology* (213) from the standpoint of general behavioristics (1950), and a *Handbook of Psychology* in 12 volumes is now being prepared under the sponsorship of the Japanese Society of Applied Psychology. A latent behavioristic influence has now become fairly well established. In 1941 Imada translated Bridgman's *Logic of Modern Physics* (33) into Japanese, and a criticism of operationism was expressed in the same year by Yatabe (209).

Experimental Psychology

Handbooks of experimental psychology have been edited by Takagi and Kido, of which Volume I (general methods of experimentation), Volume II (vision), and Volume III (audition and other sensory experience) have thus far been published (180). As to methodology, it is worth noting that interests in stochastics and factor analysis have increased among Japanese psychologists since the war, and some researches are being conducted in these areas (Koga [57], Indo [34], Iwahara [39] and others).

Perception. A book on the psy-

chology of perception was published by Ogawa, Tanaka, and Osaka (140) in 1952. Recently, Obonai (127) has surveyed western and Japanese studies on visual perception.

Experiments on perception are, in general, maintaining ties with the earlier gestalt influence, now somewhat mixed with behavioristic tendencies. Interest continues to be shown in problems of space perception generally (for example, by Akishige [9], Miyakawa [100], Osaka [148]), but problems of size and shape discrimination (as related to the perceptual constancies) seem to be particularly popular (Kume [68], Makino [77], Misumi [95]). Quantitative theories of space, as, for example, Luneburg's, are received with considerable appreciation.

The topic of figural aftereffects is being studied with enthusiasm, and in the opinion of one of us (CHG) Japanese studies in this area are being done very effectively (e.g., by Ogasawara [130], Azuma [15], Oyama [150], and others [30, 123]). Kakizaki is studying the effect of preceding conditions on retinal rivalry. Experiments on time errors have been done by Inomata (35), Nakajima (115), Ono (145), and others. Obonai's "induction theory" and Yokose's psychophysical research (207, 216) on form perception are influential contributions.

Interest in psychophysiological problems has increased since the end of the war; this is due, among other influences, to the very original work of the physiologist Motokawa (107, 108, 109) on the electrical excitability of the eye. Studies on electrical excitability are now being carried on in some psychological laboratories. They involve studies of the time course of various color effects, form distortions, contrast effects, etc.

Many areas of perception are being

examined, but it will not do at this time to describe the efforts in great detail. One trend may be worth mentioning: the increased communication of psychologists with physiologists and people in related sciences.

Audition has not become as extensive a research interest as visual perception. Nevertheless, it is interesting to observe that Wada (201) has recently (1951) published a book with the same title as Yuki's earlier volume, *A Psychology of Tone*. A recent publication concerns an analysis of Japanese vowels (110).

Other sensory systems have not been studied extensively since the war.

Animal and human learning; thinking. The greatly increased interest in animal experimentation is a major postwar development. Many experiments have been done recently at the universities of Tokyo, Kyoto, Tohoku, Hokkaido, and Osaka. The central problem has been the controversy between field theory and reinforcement theory. Studies have been made on the problems of subgoal responses (Umeoka [194] and Yagi [204]), latent learning (Nozawa [124], Asami [38]), and the effects of amount of reinforcement and amount of drive. Studies of reasoning (Suenaga [169]), place learning, and experimental neurosis (Murata [114]) are in progress at this time. It is also worth noting that a Society for the Study of Behavior Theory has been established by psychologists at the universities of Kyoto, Osaka, and Osaka City.

Considerable interest also exists in problems of human learning. Recent studies have been concerned with retroactive and proactive inhibition (Sagara, Ishiwara [37], Umemoto [193]), and researches are in progress on questions of temporal factors in interpolated materials and of generalization effects between original lists and interpolated ones. Transposition experiments on o_1 o_2 have been almost wholly restricted to the laboratory of Kwansei Gakuin University under Kotake. Studies of transposition responses in children have been performed and the developmental aspects of this type of behavior are now being examined by Sato, Okano (144), Motoyoshi (112), and others.

Little work is being performed on motor skills.

Recently a book on the psychology of learning was published by Yagi, Umeoka, and Maeda (205).

In the period 1946 to 1949, Yatabe published *A History of Thinking* (211) and three volumes of his *Psychology of Thinking* (212). Volume I of the latter work is on concept and meaning, Volume II on relations and reasoning, and Volume III on thinking of animals. These books by Yatabe, together with his *History of the Psychology of Will* (210), are authoritative and comprehensive. They are not as well known outside of Japan as they should be.

Personality. Research on personality per se has not expanded after the war to the same degree as some other areas of interest. In this connection, however, it is important to observe that many investigators are at work on experiments that are closely related to this area, for example, those that deal with the influence of motives on discrimination (92, 98, 162). In fact, most Japanese experimental psychologists show considerable interest in this topic and experiments are being done that bridge the boundary between studies of personality and perception (45). Some psychologists (e.g., Kitamura and Imada) have recently considered the problem of the ego.

Recently (1951) one of us (Sato)

published *A Psychology of Personality* (161) which discusses the training procedures of Zen Buddhism against a background of western studies.

Developmental Psychology, Educational Psychology, and Measurement

The reform of the Japanese educational system, instituted by the Occupation, involved an important program of teacher education and re-education. The re-education program has placed great stress on the data of educational psychology, and, under the new system, many school teachers have been instructed in this area. If widespread knowledge of an area is an important contributor to its vitality, then educational psychology (81, 212) is very much alive in Japan. American psychologists who went to Japan during the Occupation were usually educational psychologists, and it may be expected that the results of their efforts will show tangible returns in the way of intensified activity in this area.

Recently several books (22, 212) on the developmental aspects of educational psychology have been translated into Japanese; several books have been written on child psychology with an essentially educational emphasis, for example, one by Yamashita (206); several books on the psychology of adolescence (Katsura [49], Ushijima [198]) have also appeared. Little attention has as yet been paid to problems of maturity and old age.

Experimental studies of mental development are being carried on, notably by Sonohara (167) and Nakano (118). Takemasa published two volumes on *Developmental Psychology* (181) in 1948–1950. The large public interest in educational and developmental psychology is reflected in the fact that journals in this area are edited for the general public as well as the scientific. Emphasis is placed on the problem of learning and its relation to vocational and educational guidance.

Work continues in the general areas of intelligence and personality tests in several institutions, most notably in the National Research Institute of Education. Tanaka, for years one of the dominant figures in educational measurement, still remains a leader.

An emphasis on mental hygiene characterizes postwar educational psychology, and some psychiatrists and psychologists are now working together. In Japan, the term "mental hygiene" is used in its broadest sense to apply to normal as well as abnormal individuals. In student counseling work, many problems peculiar to Japan arise in a way that may have no parallel in the United States. In particular, political problems are said to provide an "apparent" focus of maladjustment in students. In any case the student counselor has an important role to play in modern Japan.

Clinical Psychology

The status of clinical psychology in Japan has changed greatly in the postwar period. The name "clinical psychology" is now in common use and the Society of Clinical Psychology has been formed. A journal of clinical psychology (*Clinical Psychology and Educational Counseling*) has recently been established, and now, for the first time, psychoanalytic theories are becoming common subjects of discussion. Projective tests, such as the TAT and Rorschach, play important roles in diagnosis, and nondirective methods of therapy are being studied (55, 106, 168). In all of this, the influence of

American clinical psychology is felt strongly, but in addition, some truly Japanese methods, such as Morita's (104), which are based upon a synthesis of Zen doctrines and western studies, are regarded favorably. Without doubt, clinical psychology is developing rapidly as an important field of endeavor in Japan.

Industrial Psychology

Although industrial psychology in Japan has not developed to the same degree as in America, it has nevertheless been shown that a basic core of subject matter, technique, and practice did exist prior to World War II. During World War II problems of training and efficiency became serious, and a number of psychologists worked in the Army and Navy, especially in the selection of aviators. Since the war, labor problems (training and worker morale, especially) have received more attention than previously. Several prefectures have established research institutions for work in this area, and a few universities have set up chairs of industrial psychology and vocational guidance. Work in selection shows little progress, for in a land with 85 million people and disproportionately inadequate natural resources, such procedures seem indeed to be fanciful.

Social Psychology

After the war many social problems forced themselves upon the attention of psychologists. At the same time, the theories and experiments of social psychologists in America flowed into Japan and stimulated the Japanese workers. In consequence, work in social psychology is now being carried on at an accelerated pace. At the annual meeting of the Japanese Psychological Association in 1952, about 40 studies were reported by social psychologists and, if related researches are added to this total, the number becomes considerably greater.

In the seven years since the war, the topics of central interest have changed. At first, considerable effort was spent on the analysis of social and cultural phenomena (3, 32) that occurred during and after the war. Later, attention was paid to problems of group dynamics in connection with the democratization of the Japanese people. More recently, problems of social perception, communication (28), and the measurement of attitude became a central problem not only for social psychologists but for many others as well. Most recently, psychologists have become interested in systematization (Ikeuchi [31], Suenaga [170], and others). Problems have been restated, and the possibility of treating social behavior mathematically has been examined. Several research groups have been formed, for example, the Society for the Research in Behavior Mechanisms (at Tokyo), the Association for Group Dynamics (at Kyushu), and the Society for Youth Work (at Tokyo). As yet, no special journals have been established in social psychology, but a number of systematic treatises have appeared lately, notably those by Minami (94) and Shimizu (163).

OVERVIEW

Our survey points to the fact that psychology in Japan, like psychology in most countries, did not develop rapidly during the first 40 years of its existence. Circumstances improved after World War I, and after 1926 psychology advanced in a surer fashion. By 1935 it gave promise, not

only of becoming stronger but also of becoming more diversified. Unfortunately, its favorable growth was severely inhibited by the events that led to the Sino-Japanese conflict and World War II. In effect, the war period became an interval of stagnation. Following the war, Japanese psychology underwent the readjustment common to all areas of Japanese life. At the present time, it shows signs of development to a new and stronger level.

SOME OBSERVATIONS ON CURRENT ASPECTS OF JAPANESE PSYCHOLOGY

It is certain that the picture of present-day psychology in Japan differs in some details from the one that exists in America. The following discussion is aimed at acquainting the American reader with some important facets of psychological endeavor in Japan. It is not intended that the discussion will deal with "problems." Rather, it is hoped that the description of selected areas of difference between Japan and America will acquaint the western reader with some facts that are essential to an understanding of psychology in a different culture.

Training

Under the Occupation (1945–1952) a strong effort was made to remodel the educational system of Japan.[2] All aspects of education felt the impact of the attempt, and for a while the training of psychologists came under scrutiny. For purposes of this discussion we shall speak of the pre-Occupation system as the "old" system and the program suggested by Occupation authorities as the "new" system.

The old system had a number of the characteristics of the British system. Education leading to graduate work was based upon the 6-5-3-3 system: 6 years in primary school, 5 years in middle school, 3 in preparatory school, and 3 in the university. On leaving junior college (cf. p. 444) the prospective university student had to take entrance examinations which covered many fields; each university provided its own examination. If a student passed the written examination, he was called in for an oral examination and an interview. It is important to notice that university study began in the 14th school year as in England, not in the 12th school year as in America.

During his university days, a student might concentrate in psychology. The courses required in the psychology program varied somewhat from university to university, but, in general, they seem to have covered the conventional range of subject matter despite such differences as may have existed among, for example, Kyoto, Tokyo, and Keio. Kyoto had (and has) a requirement in philosophy that does not exist at Tokyo, and Keio emphasizes work in natural sciences.

Occupation authorities made an attempt, under the new system, to change the method of selecting university students for national universities. (This change constituted a small segment of a program to reconstitute the educational system into a 6-3-3-4 sequence, analogous to the American scheme: 6 years of primary school, 3 of lower secondary, 3 of upper school, and 4 of university.) It was planned under the new system that university applicants should take the same examinations at approximately the same time throughout the nation. After a student had "passed" the national aptitude test,

it was planned that he be invited to the university of his choice for special examinations prepared by that university. If the student passed the examinations, he was to be interviewed. If he did not pass, he was to be permitted to have his files sent to one more national university. Private universities planned to operate differently, but the projected pattern was similar in many respects to the one here outlined. It turns out now that ways of utilizing the national aptitude test differ among different universities. Some universities use these tests as screening devices, and some use them only to provide additional advisory material.

Under either the old or new system, a student who graduated from a university could enter upon graduate work. Under the old system, no special courses were required during the period of graduate training. Once the undergraduate courses were completed the student could apply for admission to the dean of the department and to the Faculty of Literature (under which the psychology department usually exists). Usually, as at Tokyo, entrance examinations were required only for people who came from other universities. Under these circumstances it turned out that the graduate students in a given university had usually been undergraduates in the same university. Very often the student stayed on as an assistant to the professor for many years.[3]

The new program did not say much about selection procedures, but it did make the proposal that 30 credits of course work be required in the first year of graduate study. These cred-

its were to be accumulated in the areas of experimental, clinical, social, developmental, educational, and certain optional areas. So far, the new program has not received as much support as was expected. In 1953 about one-third of entering graduate students embarked on the new course.

The main argument against the new program is that it is expensive both for the university and for the student. Under the old plan, graduate students often received salaries as assistants, and the number of assistants required in a university partially determined the number of graduate students. Under the new plan, a student who has to devote his first year to study is not likely to receive an assistant's fee.[4]

Whatever may be its fate at other strata of the educational structure, the new system has a "hard row to hoe" at the level of graduate work. For this reason, it may here be more realistic to restrict the present discussion of graduate training to the old system, unless otherwise specified.

In general, few examinations are given during a man's graduate work, and the requirements for reports on research vary from university to university. At Tokyo and Kyoto a research report must be made once a year, but at some other universities this requirement does not exist, the final thesis being the criterion for completion of the work. As had been said, the major professor decides

---

[3] Most large Japanese universities are, in fact, not freely open to graduates of other universities. Most scholars remain in the same school from their undergraduate days until they achieve professorial rank.

[4] Where the new program does exist it has often been confusingly intermixed with the old program. The result is that few students take the graduate courses required by the new program to arrive at a possibly somewhat higher level than that specified by the M.A. degree in America. In the universities of Tokyo and Kyoto, for example, the number of students entering the graduate *courses* is only four a year. This number, of course, is small in comparison with the number of students who receive the master's degree every year in America.

when a man will finally stop his training. A doctorate degree is granted, if it is granted at all, usually after many years of professional work. It is based upon a defense of the thesis.

What opinions do Japanese psychologists have about ways of improving graduate training? The following proposals were frequently encountered and seem to represent the prevailing views of the Seminar participants: (a) establish a program leading to a terminal doctor's degree obtainable after a reasonable period of graduate study;[5] (b) differentiate departmental offerings into social, applied, and experimental psychology; (c) appoint more professors in more diverse areas of psychology;[6] (d) improve the general level of undergraduate instruction; (e) specify formal prerequisites for graduate training in each area of psychology; (f) provide more adequately supervised work in the laboratory and in the clinic; (g) improve facilities for research and graduate training; (h) place psychology in some other faculty than the Faculty of Literature.

Some of the proposals may be carried out. It is possible, for example, that psychology may vacate its place in the Faculty of Literature. Other members of the Faculty even now accept the fact that psychology is misplaced. Certain government officials seem to favor the reclassification of psychology, but others do not, probably on the grounds that, as a recognized laboratory and clinical discipline, it will become more expensive.

While external agents are helping determine the position of psychology, psychologists themselves can do a great deal in the way of advancing training procedures. They can, for example, agree on well-chosen policies concerning the effective use of present facilities. (Facilities are, in fact, adequate for many purposes, and much excellent work can be produced with them. In a few universities [e.g., Kwansei Gakuin] the equipment used in prevailing research is good.)[7] Above all, establishing policies aimed at producing broadly trained psychologists, who are skillful in the use of research techniques, will constitute an important step.[8]

Professional Opportunities

The lot of a psychologist in Japan has many desirable features, but, financially, it is no more (but probably no less) attractive (in view of prevailing economic conditions) than that of many academic people in other lands. A full professor may earn $70 to $100 per month; an assistant professor, $50 to $90; an instruc-

---

[5] In Japan the M.D. degree is given after several years of research. "Bungakuhakase" (literally, Doctor of Literature, corresponding to the western Ph.D.) is not usually given until a psychologist approaches his fiftieth birthday.

[6] Two professorships in psychology exist in each of the two most recently founded universities, Hokkaido and Osaka. Tokyo University has only recently obtained a second professorship. The universities of Kyoto, Tohoku, and Kyushu have only one chair. This situation is to be contrasted with the one that exists in philosophy. At Kyoto, for instance, there are seven chairs of philosophy. This ratio of philosophy professors to psychology professors has remained unchanged since 1906.

[7] Laboratory equipment was, in general, not obtainable during the war. In consequence much of the equipment, even in the relatively new laboratories of Tohoku and Kyushu universities, is old. The provision of funds for laboratory equipment is not yet adequate.

[8] See, for general background information, an article by M. Imada: Recent psychological thinking in Japan. *Ann. Proc. Dept. Psychol. Kwansei Gakuin Univer.*, 1954, 1, 1-9.

tor, $20 to $85; and an assistant about $20. (The salaries quoted include appropriate allowance for housing, etc.) The few jobs available in education, government, and industry probably provide slightly higher maximum salaries than the academic positions.

A main flaw in the professional picture lies in the fact that professional psychology is relatively undeveloped in Japan. Psychology has not been exploited by government, education, and industry, as has been the case in the United States, and few positions exist for psychologists outside of universities. Until, through the years, a demand is felt for psychological work in other places than psychology departments, it is unlikely that professional opportunities will expand greatly. It now remains for psychologists to exploit circumstances in which the fruitfulness of psychological methods may be shown to advantage.

Problems of Communication

Psychologists in Japan feel isolated, and, in fact, they have been isolated, not only from western psychologists but, to a greater degree than they desire, from each other.

The isolation from western psychologists is attributable to a great many factors, among which geographic position plays a minor role. The lack of contact with the western world that existed from 1937 to 1945 had its undesirable effects, and the Occupation years (1945-1952) that followed constituted a period of readjustment that did not provide the best circumstances for scientific give-and-take. (During the Occupation a number of American educational and military psychologists worked with Occupation officials, and, in fact, did establish good contacts with Japa-

nese workers. However, these pleasant relations remained indirect in most cases, because the Americans usually worked in administrative positions. It is to be hoped that more opportunities for close personal relations in the seminar and laboratory, such as existed during the Kyoto Seminar in Experimental Psychology, will be established soon. It is encouraging to observe that the appointment of George M. Haslerud as Fulbright Professor at Kyoto and the visit of Koiti Motokawa to the United States during 1953–1954 signify steps in this direction.)

The barrier of language has, as much as any other factor, served to confirm the isolation of the Japanese. Despite the fact that Japanese training in foreign languages is good, the training of western workers in Japanese is bad. It has been the lot of Japanese psychologists since their beginnings to do research which remains little known to the outside world. In consequence, Japanese psychologists have had to depend on the approval of their fellow countrymen. This situation may certainly have been attended by undesirable effects, which, fortunately, now seem to be disappearing.

Certain devices for increasing communication might be tentatively recommended without commitment as to their ultimate value. For example, Japanese psychologists could be encouraged to publish in other languages than their own. (It is worth observing in this connection that, even now, Japanese psychologists, feeling a need to converse more effectively with their western colleagues, usually add extensive English abstracts to their research reports.) Other devices such as intercultural seminars, exchange scholarships, exchange professorships, inter-

cultural institutions, etc. might pro-
duce more intimate scientific rela-
tions.   Certainly such programs
should be encouraged.

During the war and for some time
thereafter, Japanese psychologists
had little contact with each other.
The Japanese Psychological Associa-
tion may well ask itself what meth-
ods may be used to increase com-
munication among its members. The
planning of symposia on topics of
general interest might be suggested
as an area in which the Association
can make effective contributions, and
there are indications that action will
be taken along this line. Finally, it
is worth observing that devices that
improve intercultural relations can
also improve contacts among the
Japanese. One of the chief merits of
the Kyoto Seminar lay in the fact

that it attracted psychologists from
widely separated areas of Japan and
provided an appropriate focus for dis-
cussion and mutual acquaintance.

SUMMARY

An account of Japanese psychology
is presented, first from a historical
point of view, and second from the
point of view of analyzing some of its
modern aspects.

The historical sequence is broken
up into three parts: from the begin-
nings of psychology (about 1880 un-
til 1926, from 1926 through World
War II, and from the end of World
War II to the present. Some current
aspects manifested by Japanese psy-
chology are considered. The discus-
sion of these matters centers about
the topics of training, professional
opportunities, and communication.

REFERENCES

(Titles in brackets are English translations of Japanese titles of those works which have
no western abstracts. Page references in brackets refer to Japanese originals.)

1. ABBE, M. Der räumliche Effekt auf die
   Zeitauffassung. I, II. *Jap. J. exp.
   Psychol.*, 1936, **3**, 1–8 [1–52]; 1937,
   4, 1–2 [1–12].
2. ABBE, M. [Psychological correlation of
   time and space.] Nagoya: Reimeisho-
   bo, 1949.
3. ABE, J. The problem of culture in social
   psychology. *Jap. J. Psychol.*, 1949,
   20, 19–21 [13–19].
4. ABE, K. An investigation of the law of
   memory trace deviation. *Jap. J.
   Psychol.*, 1951, **21**, No. 1, 45–46 [33–
   45]; No. 2, 36–37 [29–36].
5. ABE, S. Neue Untersuchung über die
   absoluten Eindrücke im Gebiete der
   wahrnehmbaren Zeit. I, II. *Jap. J.
   Psychol.*, 1933, **8**, 35–72, 243–280.
6. ABE, S. [Studies on the racial characters
   of the races in Manchuria.] *Rep. Res.
   Inst. Kenkoku Univer.*, 1943, No. 4,
   105–180.
7. AKAMATSU, P., UTIDA, Y., & TOGAWA, Y.
   A study on schizothymics. *Jap. J.
   Psychol.*, 1940, **10**, 93–94 [971–984].
8. AKISHIGE, Y. Effekt der Entfernung
   des Darbietungsortes der Normal-
   grösse auf den Grad der Grössenkon-

stanz. *Mitt. Jurist. -Lit. Fak. Kyushu
Universität*, 1937, **4**, 37–58.
9. AKISHIGE, Y. Experimental researches
   on the structure of the perceptual
   space. *Kyushu Psychol. Stud.*, 1951,
   No. 1, 117–136.
10. AOKI, S. [An experimental study relat-
    ing to the whole and part methods of
    learning.]   In Y. Kuwata, (Ed.),
    *Essays on psychology and art in honor
    of Matsumoto Matatoro.*   Vol. I.
    Tokyo: Kaizosha, 1931. Pp. 653–667.
11. AMANO, T. [Perception of visual forms
    and its reproduction.] *Jap. J. Psy-
    chol.*, 1930, **5**, 654–674, 847–889;
    1931, **6**, 369–400.
12. AMANO, T. [Studies of the process of
    comparison.] Tokyo: Sanseido, 1944.
13. AWAJI, Y. [Army mental test.] *Jap. J.
    Psychol.*, 1926, **1**, 173–278.
14. AWAJI, Y., & OKABE, Y. A version test
    and a version coefficient. *Jap. J.
    Psychol.*, 1932, **7**, 1–54, 373–414; 1933,
    8, 417–438.
15. AZUMA, H. The effect of experience on
    the amount of Müller-Lyer illusion.
    Jap. J. Psychol., 1952, **22**, 122–123
    [111–122].

16. Chiba, T. Über die Asymmetrie der Unterschiedempfindlichkeit. *Z. Psychol.*, 1923, 92, 177–226.

17. Chiba, T. Über das "Eigenbewusstsein." *Jap. J. Psychol.*, 1927, 2, 809–814.

18. Chiwa, H. [Fatigue and practice in mental work.] Tokyo: Shinrigakukenkyukai, 1917.

19. Endo, O. [Characterology.] Tokyo: Kobundo, 1941.

20. Fujiwara, K., & Obonai, T. The quantitative analysis of figural aftereffect. II. *Jap. J. Psychol.*, 1953, 24, 177–178 [114–126].

21. Fukutomi, I. [On the delta movement.] *Rep. of II. Meeting of Jap. Psychol. Ass.*, 1929, 138–145.

22. Goto, I. (Trans.) [Goodenough, F. *Developmental psychology.*] Tokyo: Kanekoshobo, 1950.

23. Hall, G. S., & Motora, Y. Dermal sensitivity to gradual pressure changes. *Amer. J. Psychol.*, 1887, 1, 72–98.

24. Hatano, K. [Child psychology.] (Rev. Ed.) Tokyo: Dobunkwan, 1935.

25. Hayami, H., & Miya, K. Die Gestaltauffassung in der Müller-Lyerschen Figur. *Jap. J. Psychol.*, 1937, 12, 49–51 [525–552].

26. Hayashi, T. [Methodology of conditioned reflex.] Tokyo: Mikasashobo, 1940.

27. Hisata, T. Experimentelle Untersuchungen über die "Scheinbewegungen" im akustischen Gebiet. *Jap. J. Psychol.*, 1934, 9, 25–26 [367–400].

28. Hirota, K. Group problem-solving and communication. *Jap. J. Psychol.*, 1953, 24, 176 [105–113].

29. Ibukiyama, T. Experimentelle Untersuchung über Sehgrössenkonstanz. *Jap. J. Psychol.*, 1933, 8, 8 [21–34].

30. Ikeda, Hisako, & Obonai, T. The quantitative analysis of figural aftereffect. I. *Jap. J. Psychol.*, 1953, 23 [246–260]; 1953, 24, 84–85 [59–66].

31. Ikeuchi, H. A treatise on the foundation of social psychology. *Jap. J. Psychol.*, 1949, 20, 28–29 [22–28].

32. Imada, M. [Religious psychology.] Tokyo: Bunsendo, 1947.

33. Imada, M., & Ishibashi, S. (Trans.) [Bridgman, P. W. *The logic of modern physics.*] Tokyo: Sogensha, 1941.

34. Indo, T. [The phi function of gamma.] *Jap. J. Psychol.*, 1949, 19, 219–229.

35. Inomata, S. The time error in vision and audition. *Univer. Kyoto Stud. Psychol.*, 1949, 1, 6–7.

36. Iritani, C. [Group mind.] Tokyo: Daibunken, 1928.

37. Ishiwara, I. [The third factor hypothesis of retroactive inhibition.] *Jap. J. Psychol.*, 1950, 20, Suppl.

38. Iwahara, S., Asami, Chizuko, Okano, T., & Shibuya, K. A study of latent learning in rats. *Annu. Anim. Psychol.*, 1949, 3, 65 [61–64].

39. Iwahara, S. [Stochastics for the students of psychology and education.] Tokyo: Sekaisha, 1952.

40. Iwai, K. [Über Ganzheitspsychologie.] In Iwanmi (Ed.), *Handbook of educational research.* Vol. 15. 1932. Pp. 1–16.

41. Iwai, K., & Volkelt, H. Umgang des Kindes mit verschieden geformten Körpern im 9 bis 12 Monat. *Bericht. ü.d. XII. Kong. Deutsch. Gesell. f. Psychol.*, 1932, 12, 356–358.

42. Iwai, K., Sonohara, T., & Tadera, A. Studies on the tests for Japanese young children—results of the Ch. Bühler's tests given to Japanese babies 1 aged 2–7 months. *Jap. J. Psychol.*, 1935, 10, 91 [941–957].

43. Kakise, H. A preliminary experimental study of the conscious concomitants of understanding. *Amer. J. Psychol.*, 1911, 22, 14–64.

44. Kakizaki, S. The effect of preceding conditions upon binocular rivalry. *Jap. J. Psychol.*, 1950, 20, No. 2, 31–32 [24–31]; No. 4, 16–17 [11–16].

45. Kakizaki, S. [Perception and need.] *Studies in the Humanities* (Osaka City University), 1953, 4, 150–173.

46. Kamitake, M. [A conditional genetic study of leadership in group of children.] *Jap. J. Psychol.*, 1941, 16, 352–380; 1942, 17, 93–132, 296–320.

47. Kaneko, H., Ojima, S., & Miya, K. [On the friendship in the classmate of school children.] *Jap. J. Psychol.*, 1931, 6, 223–247; 1932, 7, 133–146, 429–453.

48. Kato, Masahide. An experimental genetic study of behavior forms in "the ball and field test." *Jap. J. exp. Psychol.*, 1935, 2, 9–10 [59–88].

49. Katsura, H. [Psychology of adolescence.] Tokyo: Kanekoshobo, 1950.

50. Kido, M. [Weltanschauungen der alten Japaner.] Tokyo: Iwanami, 1930.

51. Kido, M. [General outline of psychology.] Tokyo: Iwanami, 1931.

52. KIRIHARA, S. [Menstruation and work ability.] Tokyo: Toyoshokan, 1943.

53. KIRIHARA, S. [Mental measurement.] Tokyo: Sanseido, 1944.

54. KITAMURA, S. Untersuchung über die Typen des Vorstellungslebens in Bezug auf das Icherlebnis. I. *Tohoku Psychol. Folia*, 1939, **7**, 189–208; II, 1940, **8**, 1–30.

55. KODAMA, A. [Methods of diagnosing personality. II.] In. A. Kodama, *et al.* (Eds.), *Handbook of psychology.*, Vol. 7. Tokyo: Nakayamashoten, 1953.

56. KOGA, Y. Measurement of attitudes towards sport. *Jap. J. Psychol.*, 1934, **9**, 66–67 [953–977].

57. KOGA, Y. [Type-factors in aesthetic appreciations and personality traits. A study of inverted factor technique.] *Jap. J. appl. Psychol.*, 1939, **5**, 1–46.

58. KOGA, Y., & MORANT, G. M. On the degree of association between reaction times in the case of different senses. *Biometrika*, 1923, **15**, 346–372.

59. KOHRA, T. [Characterology.] (Rev. Ed.) Tokyo: Sanseido, 1938.

60. KOTAKE, Y. Psycho-physiological approach to the conditioned reflex in the human subject. *Jap. J. Psychol.*, 1951, **21**, 16–17 [1–16].

61. KOTAKE, Y., & TAGAWA, K. On the delay of conditioned GSR in man. *Jap. J. Psychol.*, 1952, **22**, 5–6 [1–5].

62. KOTAKE, Y., & MIHAMA, H. Conditioning of pupillary reflex in man. *Jap. J. Psychol.*, 1952, **22**, 88 [79–87].

63. KUBO, S. Über die phänomenale Veränderungen der optischen Scheinbewegungen, die in den Rhythmusgestalt eingebettet ist. *Jap. J. Psychol.*, 1931, **6**, 477–541.

64. KUBO, YOSHIHIDE. [The national intelligence tests.] Tokyo: Jidokenkyujokiyo, 1922.

65. KUBO, YOSHIHIDE. [Handbook of experimental psychology.] Tokyo: Chunbunkwan, 1926, 1927.

66. KUBO, YOSHIHIDE. [Child Psychology.] Tokyo: Fujiishoten, 1941.

67. KUBO, YOSHITOSHI. Experimental studies of so-called "form constancy phenomenon." *Jap. J. Psychol.*, 1936, **11**, 22 [265–278].

68. KUME, KYOKO. [On the limitation of the applicability of Brunswik-Thouless' index of the size constancy.] *Jap. J. Psychol.*, 1948, **19**, 70–82.

69. KURAISHI, S. An experiment on the controlled association. A psychological study of thinking. I. *Jap. J.*

70. KURODA, G. [Objective conditions of the binocular predominance, rivalry and mixture.] *Jap. J. Psychol.*, 1919, **1**, 47–74.

71. KURODA, G. Zur Grenzbestimmung der binocularen Phänomene. *Psychol. Forsch.*, 1925, **6**, 282–297.

72. KURODA, R. Experimental researches upon the sense of hearing in lower vertebrates, including reptiles, amphibians and fishes. *Comp. Psychol. Monogr.*, 1926, **3**, No. 16.

73. KURODA, R. Stereopsychology: its scope and method. *Acta psychol. Keijo*, 1931, **1**, 69–82.

74. KURODA, R. [Animal psychology.] Tokyo: Sanseido, 1933.

75. KUWATA, Y. [Folk psychology of Wundt.] Tokyo: Kaizosha, 1924.

76. MAEDA, Y. [A contribution to the analysis of the reproduction mechanism. I, II.] *Jap. J. Psychol.*, 1940, **15**, 181–206; 261–281.

77. MAKINO, T. The method of investigating the constancy of shape. *Jap. J. Psychol.*, 1950, **20**, 12–13 [1–12].

78. MANABE, H. [On the relation between the brightness and light constellation in the visual field.] *Jap. J. Psychol.*, 1939, **14**, 1–8.

79. MARUI, K. Psychoanalytische Studie über einen Fall hysterischer Amaurose. *Arbeit. Psychiat. Instit. d. K. Japan. Univer. zu Sendai*, 1932, **1** 68–70 [1–24.]

80. MASAKI, M. Die experimentelle Untersuchung au den Anschauungsbildern. *Jap. J. Psychol.*, 1930, **5**, 7–10 [355–401.]

81. MASAKI, M. [Fundamental problems of educational psychology.] Tokyo: Dogakusha, 1950.

82. MASAKI, M., & YODA, A. [Psychology of character.] Tokyo: Tokoshoten, 1937. (Rev. 1941.)

83. MASUDA, K. [Comparative psychological study of will, including experiments of packing of new hatched chicken and puzzle box experiment with some birds.] *Tetsugaku-Zassi*, 1908–1909, **23**, 950–970, 1029–1070, 1139–1176, 1270–1302; **24**, 36–60, 239–262, 352–374, 552–588.

84. MASUDA, K. [An introduction to experimental psychology.] Tokyo: Shibundo, 1926.

85. MASUDA, K. [Methodology of psychology.] Tokyo: Iwanami, 1934.

86. MATSUBAYASHI, A. Forschung über die

Tiefenwahrnehmung. I–X, *Mitteilung Acta ophthalm. Japonicae*, 1937–1938, **41**, 94–95, 150–151, 158, 167 [1289–1302, 2055–2074, 2151–2161, 2257–2268], **42**, 1, 15, 26–27, 31–32, 82, 133 [2–21, 230–241, 366–377, 480–491, 1920–1929, 1185–1196]

87. MATSUMOTO, M. Researches on acoustic space. *Stud. Yale Psychol. Lab.*, 1897, **5**, 75 p.

88. MATSUMOTO, M. Ten lectures on experimental psychology. Tokyo: Kodokwan, 1914.

89. MATSUMOTO, M. [Psychocinematics.] Tokyo: Rikugokwan, 1914.

90. MATSUMOTO, M. [Psychology of intelligence. Tokyo: Kaizosha, 1925.

91. MATSUMOTO, M. ["Preface"] to the Report of IV Meeting of Jap. Psychol. Ass., 1933.

92. MATSUMURA, K. [On the process of satiation. A study of character.] *Jap. J. Psychol.*, 1942, **17**, 38–73.

93. MINAMI, H. [An experimental study on the determining factors of visual space perception.] *Jap. J. Psychol.*, 1940, **15**, 153–180.

94. MINAMI, H. [Social psychology.] Tokyo: Kobunsha, 1949.

95. MISUMI, J. Experimental study of the development of visual size constancy in early infancy. *Jap. J. Psychol.*, 1950, **20**, 25–26 [16–25].

96. MITUI, T. Experimental investigation on the process of speculation. *Jap. J. Psychol.*, 1928, **3**, 30–32 [769–796].

97. MIYA, K. Die Gruppenstruktur des Haushuhns. *Jap. J. Psychol.*, 1936, **11**, 3–4 [20–37].

98. MIYAGI, O. [The prepsychotic personality of schizophrenia.] *Jap. J. Psychol.*, 1948, **19**, 117–128.

99. MIYAGI, O. [Dreams.] Tokyo: Iwanami, 1953.

100. MIYAKAWA, T. Experimental research on the structure of visual space when we bend forward and look backward between the spread legs. *Jap. J. Psychol.*, 1950, **20**, 22–23 [14–22].

101. MIYOSHI, M. [An experimental and theoretical study of fluctuation problems. I, II.] *Jap. J. Psychol.*, 1940, **15**, 282–304; 1943, **18**, 45–64.

102. MORINAGA, S. [Einige Überlegungen über die Müller-Lyersche Figur.] *Jap. J. Psychol.*, 1941, **16**, 26–39.

103. MORINAGA, S., & OGASAWARA, J. Anomalies in the visual perception by the lesion of the brain. *Jap. J. Psychol.*, 1949, **19**, 184–190.

104. MORITA, S. [Nature and therapy of "nervousness."] Tokyo: Tohodo, 1928.

105. MORIYA, M. [Psychology of early childhood.] Kyoto: Naigaishuppan, 1944.

106. MOTOAKI, H. [Methods of diagnosing personality. I.] In A. Kodama *et. al.*, (Eds.), *Handbook of psychology.*, Vol. 7. Tokyo: Nakayamashoten, 1953.

107. MOTOKAWA, K. Retinal processes and their role in color vision. *J. Neurophysiol.*, 1949, **12**, 291–303.

108. MOTOKAWA, K. The field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413–426.

109. MOTOKAWA, K., & EBE, M. The physiological mechanism of apparent movement. *J. exp. Psychol.*, 1953, **45**, 378–386.

110. MOTOMIYA, Y. On the characteristic frequency bands of the Japanese vowels. *Jap. J. Psychol.*, 1935, **10**, 78–79 [831–837].

111. MOTORA, Y. [Outline of systematic psychology.] Tokyo: Hobunkwan, 1915.

112. MOTOYOSHI, R. "Situation" or "attitude" in selective response. *Univer. Kyoto Stud. Psychol.*, 1949, **1**, 4–5.

113. MUKUNO, K. [Apparent movement on the skin.] *Jap. J. Psychol.*, 1932, **7**, 695–734.

114. MURATA, K. Maze learning and audiogenic seizure in the rat. *Univer. Kyoto Stud. Psychol.*, 1949, **1**, 22–24.

115. NAKAJIMA, S. On the time error in successive comparison of time. *Jap. J. Psychol.*, 1951, **21**, No. 3–4, 44–45 [36–44].

116. NAKAJIMA, T. Contributions to the study of the affective processes. *Amer. J. Psychol.*, 1909, **20**, 157–193.

117. NAKAMURA, K. [A study on the experimental method of thinking process.] *Jap. J. Psychol.*, 1944, **19**, 1–11.

118. NAKANO, S. [Thinking in children.] Tokyo: Kanekoshobo, 1949.

119. NARASAKI, A. [Psychodynamic studies of children and youth.] Tokyo: Chubunkwan, 1922.

120. NARUSE, G., & OBONAI, T. Decomposition and fusion of mental images in a drowsy or post-hypnotic hallucinatory state. *Jap. J. Psychol.*, 1952, **22**, 188 [175–188].

121. NISHI, T. [On the apparent size of the sun and the moon.] In Y. Kuwata (Ed.), *Essays on psychology and art.* Vol. I. Tokyo: Kaizosha, 1931. Pp. 301–328.

122. NOGAMI, T. [Psychology and education

of adolescents.] Tokyo: Dobunshoin, 1937.

123. NOZAWA, S. Prolonged inspection of figure and its after-effect. I. *Jap. J. Psychol.*, 1953, 24, 85–87 [1953, 23, 217–234; II. 1953, 24, 47–58].

124. NOZAWA, KAZUKA. [An experimental study on latent learning.] *Jap. J. Psychol.*, 1951, 21, 103.

125. OBONAI, T. Contribution to the study of psychophysical induction. I. Experiments on the illusion of contrast and confluence. *Jap. J. Psychol.*, 1933, 8, 4–7 [1–20].

126. OBONAI, T. [Induction theory.] In K. Tanaka & M. Kido (Eds.), *Essays on psychology in memory of 77th year of M. Matsumoto.* 1943, 130–173.

127. OBONAI, T. [Visual perception.] In A. Kodama *et al.* (Eds.), *Handbook of psychology.* Vol. 4. Tokyo: Nakayamashoten, 1953.

128. ODANI, S. Experimental study upon brain mechanisms in learning. I, II, III. *Jap. J. exp. Psychol.*, 1933, 1, 25–30 [143–210]; 1934, 2, 17–21 [129–165]; 1936, 3, 31–32 [277–316].

129. OGASAWARA, J. Über den Einfluss des phänomenalen Abstandes auf das Auftreten von beta (stroboskopicher) Bewegung. *Jap. J. Psychol.*, 1936, 11, 11 [109–922].

130. OGASAWARA, J. Untersuchungen über die Konstanzerscheinungen—I. Die Konstanz der gesehenen Geschwindigkeit. *Jap. J. Psychol.*, 1938, 13, 2–4 [29–44].

131. OGAWA, T. The phenomenological displacement of the path of visual movement. *Jap. J. Psychol.*, 1938, 13, 29–31 [307–334].

132. OGAWA, T. The induced bodily movement. *Jap. J. Psychol.*, 1951, 21, 20–21 [14–20].

133. OGUMA, T. [Psychology of dreams.] Tokyo: Etsuzando, 1918.

134. OHBA, C. [Folk psychology.] Tokyo: Kobundo, 1941.

135. OHNISHI, K., TAKAGI, M., NAGASHIMA, S., IMURA, T., INUI, T., & MOTOAKI, H. [Psychology of character.] Tokyo: Kanekoshobo, 1952.

136. OHTOMO, S. [Diagnostics of education.] Tokyo: Baifukwan. Vol. I. 1928; Vol. II. 1933.

137. OHTSUKA, N. Über die absolute und relative Wahl beim Affen *Cereopithecus sp. Acta Psychol. Keijo*, 1937, 3, 32–44.

138. OHTSUKI, K. [Experimental psychology]. Tokyo: Seibido, 1911.

139. OHWAKI, Y. [Eidetic imagery of children.] Tokyo: Tsuruhobo, 1936.

140. OHWAKI, Y. Laperimentelle Untersuchungen über die Erscheinungsweise der Vorstellungen. *Tohoku Psychol. Folia*, 1939, 7, 51–118.

141. OHWAKI, Y., & ONIJIMA, T. Function of the ground as "frame work" in the perception of size. *Tohoku Psychol. Folia*, 1951, 12, 53–66.

142. OKABE, T. An experimental study of belief. *Amer. J. Psychol.*, 1910, 20, 563–596.

143. OKAMOTO, H. [An experimental study on the colour-discrimination of a crow.] *Jap. J. Psychol.*, 1939, 14, 270–296.

144. OKANO, T. [Comparison of the simultaneous and successive discrimination of the rat.] *Jap. J. Psychol.*, 1953, 24, 132.

145. ONO, S. On the positive time error in the successive comparison of brightness. *Jap. J. Psychol.*, 1950, 20, 15 [6–15].

146. ONOSHIMA, U. Über die Abhängigkeit akustischer Intensitätsschritte von einem umfassenden Tonverband. *Psychol. Forsch.*, 1928, 11, 267–289.

147. ONOSHIMA, U. [Twelve lectures on recent psychology.] Tokyo: Baifukwan. 1930.

148. OSAKA, R. The anisotropy of visual space—on the moon illusion—*Univer Kyoto Stud. in Psychol.*, 1949, 1, 15–20.

149. OSAKA, R., OGAWA, T., & TANAKA, Y. [Psychology of perception.] Tokyo: Kanekoshobo. 1952.

150. OYAMA, T. Experimental studies of figural after-effects. I. Temporal factors. *Jap. J. Psychol.*, 1953, 23, 288–289 [239–245].

151. RIKIMARU, J. A study of "taste-blindness." *Jap. J. Psychol.*, 1934, 9, 11–12 [87–109].

152. SAGARA, M. Homogenität und Heterogenität der Prozesse und Spuren bei der Reproduktion. *Jap. J. Psychol.*, 1936, 11, 12–18 [123–144].

153. SAGARA, M. [Assimilative process as a factor of inhibition in memory.] *Jap. J. Psychol.*, 1937, 14, 9–28, 367–377.

154. SAKUMA, K. (trans.) [Köhler W. *Gestaltpsychology.*] (Rev. Ed.) Tokyo: Uchidarokakuho, 1938.

155. SAKUMA, K. [The basic consciousness as directly experienced—an introduction to "the structural varieties of consciousness."] In Y. Kuwata (Ed.), *Essays on psychology and art.* Vol. I, Tokyo: Kaizosha, 1931. Pp. 301–328.

156. SAKUMA, K. The structure of the Jap-

anese language. *Kyushu Psychol. Stud.*, 1951, No. 1, 1-89.

157. SAKUMA, K., & LEWIN, K. Die Sehrichtung monokularer und binokularer Objekte bei Bewegung und das Zustandekommen des Tiefenobjektes. *Psychol. Forsch.*, 1925, 6, 298-357.

158. SAKUMA, K., & YATABE, T. Die Rolle des Senkungswinkels in der Tiefenwahnehmung. *Jap. J. Psychol.*, 1928, 3, 1-5.

159. SATO, K. [Über Kruegers Gefühlstheorie.] *Tetsugakukenkyu*, 1931, 16, 103-136.

160. SATO, K. Studies on the apprehension of relation. Part II. 1. Experimental analysis of the problem of transfer of selective response. *Jap. J. exp. Psychol.*, 1936, 3, 27-29 [219-261].

161. SATO, K. [Psychology of personality.] (Rev. Ed.) Tokyo: Sogensha, 1953.

162. SATO, K. [Frustration.] In A. Kodama et al. (Eds.), *Handbook of psychology*, Vol. 7. Tokyo: Nakayamashoten, 1953.

163. SHIMIZU, K. [Social psychology.] Tokyo: Iwanami, 1951.

164. SHIRAI, TSUNE. [A study on the development of apprehension of relation.] *Rep. Res. Educ. Psychol.*, 1941, 1, 74-129.

165. SONOHARA, T. Systematische Beobachtung über die Tagsverläufe von fünf Neugeborenen in den ersten 10 Lebenstagen. *Jap. J. exp. Psychol.*, 1936, 3, 23-26 [175-186].

166. SONOHARA, T. [On the apprehension of depth in a two-year-old child.] *Jap. J. exp. Psychol.*, 1939, 5, 149-172.

167. SONOHARA, T. [Intelligence.] "Modern psychology." Tokyo: Kawadeshobo, 1942, 4, 64-130.

168. SOTOBAYASHI, D. [Diagnostics of characters.] Tokyo: Makishoten, 1950.

169. SUENAGA, T. [The effect of experience on reasoning in rats.] *Annu. Anim. Psychol.*, 1948, 2, 82-103.

170. SUENAGA, T. [Basic problems of social psychology.] *Psukhe*, 1948, No. 2, 13-20.

171. SUSUKITA, T. [Personality structure of the races in Manchuria.] *Rep. Res. Inst. Kenkoku Univer.*, 1943, No. 5, 105-164.

172. SUTO, Y. [On the effect of the phenomenal distance upon the time perception.] *Jap. J. Psychol.*, 1941, 16, 95-115.

173. SUZUKI, H. [Individual tests for measuring intelligence.] (Rev. Ed.) Tokyo: Toyotosho, 1948.

174. TACHIBANA, K. [Senescence.] Tokyo: Kobundo, 1941.

175. TAKAGI, S. [Fundamental problems of the study of perception.] In Y. Kuwata (Ed.), *Essays on psychology and art*. Vol. 1. Tokyo: Kaizosha, 1931. Pp. 301-328.

176. TAKAGI, S. [Problem of perception.] In Iwanami (Ed.), *Handbook of philosophy*, 1933.

177. TAKAGI, S. An experimental study of the discrimination and constancy of form in the tomtit (*Sittiparus varius varius*). *Jap. J. Psychol.*, 1933, 8, 47-48 [521-548].

178. TAKAGI, S. The influence of the immediate backgrounds upon the transposition of selective responses for brightness in the varied tit (*Sittiparus varius varius*). *Jap. J. Psychol.*, 1935, 10, 71-73 [789-804].

179. TAKAGI, S. [An experimental analysis of "the factor of proximity" and "the factor of equality."] *Jap. J. Psychol.*, 1940, 15, 1-16.

180. TAKAGI, S., & KIDO, M. (Eds.) [Handbook of experimental psychology.] Vol. I-III. Tokyo: Iwanami, 1951, 1952, 1953.

181. TAKEMASA, T. [Developmental psychology.] Vol. I-II. Tokyo: Sekaisha, 1948-1950.

182. TAMAIKE, J. On the change of the phenomenal size of figure in correspondence with the structure of the visual field. *Jap. J. Psychol.*, 1933, 8, 51 [579-587].

183. TANAKA, K. [Human engineering.] Tokyo: Yubunkan, 1922.

184. TANAKA, K. [A study of the effect of oxygen deprivation on the efficiency of mental and physical works.] *Jap. J. Psychol.*, 1925, 3, 1-57; 115-162.

185. TANAKA, K. [Manual of intelligence tests (Form B).] (Rev. Ed.) Tokyo: Fujiishoten, 1949, (1936).

186. TANAKA, Y. [Some remarks on psychophysical methods.] *Jap. J. Psychol.*, 1952, 23, 93-102.

187. TODA, M. Measurement of intuitive-probabilities by a method of game. *Jap. J. Psychol.*, 1951, 22, 39-40 [29-38].

188. TOGAWA, Y. [The structure of personality.] Tokyo: Genkosha, 1944.

189. TOMEOKA, Y. Some experimental observation on the process of doubt. *Jap. J. Psychol.*, 1925, 3, 23-28 [307-322].

190. TSUJIOKA, B. [A factor-analytic study of Kraepelin-Uchida addition test.]

Kambetsu ("Diagnosis"), 1952, No. 4, 9–18.

191. UCHIDA, Y. [A manual of Kraepelin-Uchida addition test.] Tokyo: Nihon-seishingijutsukenkyusho, 1952.

192. UENO, Y. Experimental studies on the psychogalvanic reflex. *Jap. J. Psychol.*, 1935, **10**, 84–85 [875–888].

193. UMEMOTO, T. [The relation between serial position effects and retroactive, and proactive inhibition.] *Jap. J. Psychol.*, 1951, **21**, 98.

194. UMEOKA, Y. [On "sub-goal."] *Annu. Anim. Psychol.*, 1948, **2**, 14–24.

195. UMEOKA, Y., & YOSHIMI, Y. The effect of "percentile reinforcement" upon simple relearning situation. *Jap. J. Psychol.*, 1952, **22**, 101–102 [89–101].

196. UMEZU, H. [On the study of "tactual space."] In K. Hirose (Ed.), *Essays on recent psychology in memory of Koreshige Masuda.* Tokyo: Iwanami, 1935. Pp. 126–146.

197. USHIJIMA, Y. [Experimental study of learning process.] *Jap. J. Psychol.*, 1928, **3**, 322–363, 511–524.

198. USHIJIMA, Y. [Psychology of adolescence.] (Rev. Ed.) Tokyo: Ganshodo, 1941.

199. UTSUNOMIYA, S., & NAKANISHI, N. Experiment on the distance perception in the visual movement. *Jap. J. exp. Psychol.*, 1936, **3**, 9–10 [53–62].

200. WADA, Y. The influence of tonal background upon time errors in the successive comparison of intensity of tones. *Jap. J. Psychol.*, 1935, **10**, 47–48 [391–408].

201. WADA, Y. [A psychology of tone.] Tokyo: Sogensha, 1950.

202. WATANABE, T. [Personology, the study of personality.] Tokyo: Hakubunkwan, 1912.

203. YAGI, B. The influence of form upon the "Liebman Effect." *Jap. J. Psychol.*, 1938, **13**, 21–22 [213–235].

204. YAGI, B. [A preliminary study of the effect of sub-goal on maze behavior.] *Annu. Anim. Psychol.*, 1948, **2**, 69–81.

205. YAGI, B., UMEOKA, Y., & MAEDA, Y. [Psychology of learning.] Tokyo: Kanekoshobo, 1952.

206. YAMANOUCHI, T. [Animal psychology.] Tokyo: Yokendo, 1938.

207. YAMASHITA, T. [Child psychology.] Tokyo: Kobunsha, 1947.

208. YAMAZAKI (KISHIMOTO), S. The phenomenal simultaneity in the field of visual movement. *Jap. J. Psychol.*, 1935, **10**, 53–54 [545–566].

209. YATABE, T. [A critique of operationism.] *Jap. J. Psychol.*, 1941, **18**, 310–328.

210. YATABE, T. [A history of psychology of will.] Tokyo: Baifukan, 1942.

211. YATABE, T. [A history of psychology of thinking.] Baifukan, 1948.

212. YATABE, T. [Psychology of thinking.] Tokyo: Baifukan. Vol. I, 1948, Vol. II, 1949, Vol. III, rev. 1954 (1945).

213. YATABE, T. [Introduction to psychology.] Tokyo: Sogensha, 1950.

214. YODA, A. (Ed.) [Educational psychology.] Tokyo: Kanekoshobo, 1950.

215. YODA, A., MASAKI, M., & NAGSHIMA, S. (Trans.) [Jersild, A. T. *Child development and the curriculum.*] Tokyo: Shinkyoikukyokai, 1949.

216. YOKOSE, Z., & UCHIYAMA, M. The measurement of field forces in visual perception. *Jap. J. Psychol.*, 1951, **22**, 55–56 [41–55].

217. YOKOSE, Z., & KAWAMURA, H. A study of the direction of the field force in shape perception. I. *Jap. J. Psychol.*, 1952, **23**, 201 [133–143].

218. YOKOYAMA, M. Affective tendency as conditioned by color and form. *Amer. J. Psychol.*, 1921, **32**, 81–107.

219. YOKOYAMA, M. The nature of the affective judgment in the method of paired comparisons. *Amer. J. Psychol.*, 1921, **32**, 357–369.

220. YOKOYAMA, M., & SAITO, K. [On the proactive inhibition in learning.] *Jap. J. Psychol.*, 1950, **20**, 56.

221. YUKI (HIROSE), K. [A psychology of tone.] Tokyo: Gakuseisha, 1933.

222. YUKI, K. Study of motional phenomena in auditory region. *Jap. J. Psychol.*, 1933, **8**, 67–68 [745–789].

THE LEADERLESS GROUP DISCUSSION[1]

BERNARD M. BASS
Louisiana State University

The purpose of this review is to describe the leaderless group discussion (LGD): its history, applicability, method of administration, and reliability and validity as a technique for assessing leadership potential. We shall discuss the major variants in procedure that have been tried, and indicate the effects of these variations on LGD reliability and validity. We shall not consider the initially leaderless discussion's role as a basic research tool for studying the development of leadership as it has been employed by investigators such as Carter (29, 30, 31, 32), Pepinsky, Siegel, and Vanatta (61), and Bell and French (24), although we will make use of their findings wherever they shed light on the LGD as an assessment device.

HISTORY OF THE LEADERLESS GROUP DISCUSSION

According to Ansbacher (2), the originator of the method was J. B. Rieffert, who directed German military psychology from 1920 to 1931. The technique, called by first users the *Schlusskolloquium* or *Rundgespräch*, was aimed at showing behavior "toward equal partners." The German Army used the procedure until about 1939, while the Navy continued employing it in their selection programs until late in World War II. Various German civilian agencies now appear to be employing the LGD as an assessment tool.

Influenced by the German developments in situational testing, by 1942 the British War Selection Board introduced such tests into their battery for selecting Army officer candidates. The basic series of leaderless group tests was evolved by Bion and included the LGD (41, 45, 70). A similar program was established by the British Navy (70).

At the end of the war, the LGD was employed by Fraser (39, 40) as a device for screening British management trainees, and by Vernon (68, 69) for testing British Civil Service applicants.

Similar developments took place in Australia (43, 65), South Africa (3), and Norway.[2]

The OSS Assessment Staff (59) appears to have initiated use of the LGD in the United States late in World War II. American federal and state civil service examiners began trying out the technique at the end of the war (5, 26, 27, 53).

Approximately 25 per cent of the 190 Civil Service agencies surveyed by Fields (37) reported using the LGD. A Federal Civil Service manual appeared in 1952 (56). The LGD has also been employed by several American industrial and business firms. Its rapid acceptance has led both Meyer (58) and Douglas (35) to caution about overestimating its validity and utility.

---

[1] Several parts of this article were summarized for presentation at a Symposium on Situational Performance Tests, Sixty-first Annual Meeting of the American Psychological Association, Cleveland, Ohio, September 7, 1953.

[2] Private correspondence from V. C. Jahl, Chief Psychologist, Norwegian Armed Forces.

We attribute the use of the LGD primarily to its relative ease of administration compared to individual interviews, especially where large numbers of applicants are involved, as well as to its face validity compared to paper-and-pencil tests.

It may be that the face validity of the LGD really is what Gulliksen (44) has labeled *intrinsic* validity. A number of *extrinsic* validity studies are now available which may justify ex post facto the possibly widespread premature acceptance of the LGD as a valid measure of the tendency to display successful leadership.

Several investigators have assumed that the LGD is intrinsically valid as a means of appraising successful leadership behavior. Pepinsky, Siegel, and Vanatta (61) have used LGD performance as a criterion for evaluating the effects of counseling; Bell and French (24) and Carter and his associates (29, 30, 31, 32) have assumed that they are studying leader behavior when observing discussions.

Another reason for the continuing interest in employing situational tests to forecast or appraise leadership behavior may be due to the nature of leadership and psychometrics.

In reaction to the earlier emphasis on the effects of individual differences on leadership behavior, there has been, more recently, an emphasis on situational effects on leader behavior. That both sources of variance are important for leadership behavior can be accounted for fully only after analyzing the main effects and the interaction effects of situational differences and individual differences in motivation, behavioral history, and biological level (maturity, heredity, and integrity of the CNS).[3]

[3] This formulation was originally proposed by W. P. Hurder.

It follows that any test designed to forecast leadership potential and intended to have some generality of application would need to: (*a*) share elements in common with a variety of situations; (*b*) vary directly with the situation for which predictions are to be made.

Most psychometric test procedures have been developed to meet only the first requirement, for many behaviors are relatively little influenced by situational changes. Thus, one's speed-of-arm movement or spatial visualization accuracy remains fairly constant over a wide range of situations. On the other hand, since the effects on leadership behavior of situational change are large, psychometric methods like the LGD, and other situational tests which meet both requirements, would appear to offer more promise initially as psychometric techniques for forecasting leadership behavior.

With reference to the first requirement, the LGD, in common with a range of other situations in which leader behavior is to be appraised or predicted, appears to share such elements as the need for would-be leaders: to communicate effectively, to overcome inertia, to solve various interaction problems, to meet deadlines, and to reach consensus.

With reference to the second requirement, the LGD and other situational tests, by their very construction and administration, tend to vary consistently with the nature of the examinees and the real-life situations for which personnel are being chosen. Candidates for positions of leadership are assessed among administrators by observing them solving abstracts of *administrative* problems *among administrative trainees*. Tests of the same individuals for positions as Army officers would involve

studying them solving *military* problems during interaction with *officer candidates*. To some extent, therefore, situational tests may be able to take situational variations into account.

APPLICABILITY OF THE LEADERLESS GROUP DISCUSSION

According to published reports, the LGD has been used to assess candidates for many professions and occupations. The examinees have included: officer candidates (**2, 41, 43, 45, 70, 71**); OSS agent applicants (**59**); advanced Naval (**31, 32**), Air Force, and Army ROTC cadets (**15**); industrial management trainees (**39, 40, 65**); industrial executives and supervisors (**21, 22**); shipyard foremen (**53**); supervisory labor mediators (**42**); civil service supervisors and administrators (**3, 54, 68, 69**); applicants for foreign service (**68, 69**); graduate engineer trainee applicants (**9, 10**); sales trainee applicants (**9, 10**); public health physicians (**5, 26, 27**); teachers (**47**); visiting teachers (**51**); and social service workers (**52**).

DESCRIPTION OF THE LEADERLESS GROUP DISCUSSION

The basic scheme of the LGD is to ask a group of examinees, as a group, to carry on a discussion for a given period of time. No one is appointed leader. The examiners do not participate in the discussion, but remain free to observe and rate the performance of each examinee. To date, there has been little standardization by all examiners of the number of discussants per group, the length of testing time, type of problems, if any, presented to the candidates, and the directions given to them. Also, the number of raters has varied, as has the seating arrangement. Examiners have differed in the kinds of behavior they have observed and rated and in the extent to which their ratings have been attempts to describe the behavior they have observed, rather then attempts to make inferences about the personality of the candidates.

Factors Rated by Observers

Unless otherwise indicated, the LGD results we shall describe are either based on observers' ratings of the amount of *successful* leadership displayed[4] in the discussion, or are inferences about the personalities of the candidates based on observations of this behavior.

In regard to personality inferences, Couch and Carter (**34**), following a factorial analysis of observers' inferences based on several kinds of situational tests including the LGD, found that three independent factors could be isolated which accounted for situational test ratings. The factors and the ratings with highest loadings on each factor were: (*a*) Individual Prominence (authoritarianism, confidence, aggressiveness, leadership, striving for recognition); (*b*) Group Goal Facilitation (efficiency, cooperation, adaptability, pointed toward group solution); (*c*) Group

---

[4] A leadership act is said to occur when member A of a group behaves in a way directed toward changing another member, B's, behavior. More specifically, a leadership act occurs when A's behavior is directed toward: (*a*) changing the intensity and/or direction of B's motivation, and/or (*b*) restructuring B's abilities to cope with the situation and reduce B's needs (**13**).

All *attempted* leadership acts in which A reaches his goal of changing B are considered *successful leadership*. If B's change in behavior brought about by A's successful leadership leads to need-satisfaction for B and for A (apart from A's satisfaction in being a successful leader), A's successful leadership act is considered effective (**46**).

Sociability (sociability, adaptability, pointed toward group acceptance).

As will be pointed out later, many of these specific inferences by observers, such as those concerning authoritarian tendencies, may be valid only as descriptions of performance in leaderless groups and actually may be inversely related to attitudes and performance in real life.[5]

A similar factorial analysis of OSS situational test data by Sakoda (62) uncovered three factors which resemble Couch and Carter's: (a) Physical Energy (energy and initiative, physical ability, leadership); (b) Intelligence (effective intelligence, observing and reporting, propaganda skills); (c) Social Adjustment (social relations, emotional stability, "security"). Carter (29) noted that similar factors appeared in several studies when real-life leader behavior was rated and factor analyzed.

In aiming to estimate successful leadership behavior, the descriptive check lists of behavior in leaderless discussions that evolved from various studies made by the author and associates appear to concern primarily what Hemphill (46) has labeled "Initiation of Structure," a factor which has emerged in several of the Ohio State Leadership Studies (e.g., 38) and which has similarities to Couch and Carter's "Group Goal Facilitation."

In the check list used in most studies by the author and his co-workers to assess leader behavior, raters are asked to indicate whether each candidate showed the following behaviors "a great deal" (4 pts.), "fairly much" (3 pts.), "to some degree" (2 pts.), "comparatively little" (1 pt.), or "not at all" (0 pts.): (a) Showed initi-

ative; (b) was effective in saying what he wanted to say; (c) clearly defined or outlined the problems; (d) motivated others to participate; (e) influenced the other participants; (f) offered good solutions to the problem; (g) led the discussion. The sum of ratings received on all seven items is the examinee's performance score.

In an unpublished study by Pruitt and the author, seven items were rated which aimed at assessing the extent to which each LGD participant showed "consideration for the welfare of his associates"—a factor of leader behavior uncovered in various Ohio State Leadership Studies (38). This factor is similar to Couch and Carter's Group Sociability and Sakoda's Social Adjustment factors. The seven items included: (a) Engaged in friendly jokes and comments; (b) made others feel at ease; (c) complimented others; (d) helped others; (e) encouraged others to express their ideas and opinions; (f) had others share in making decisions with him; (g) helped settle conflicts. At the same time, a list of seven "Initiation" behaviors were observed and rated.

The intercorrelation between 80 LGD participants' Initiation and Consideration scores was .78, too high to warrant continued use of both assessments. More significantly, the mean rating on any single Initiation behavior was 2.2 points, while the mean rating on any single Consideration item was 0.75 points. Thus, although raters could assess reliably both types of leader behavior ($r_{tt} = .90, .85$), much less Consideration behavior appeared, and most of it was exhibited by Initiators.

On the basis of this, and evidence to be presented later, we suspect that the LGD rating is more an assessment of the tendency to initiate structure in an initially unstructured

[5] For a discussion of leadership and authoritarianism, the reader is referred to Hollander (48).

social situation—one of several types of successful leadership behavior—than an assessment of tendencies to be considerate.

Conditions Affecting LGD Leadership Ratings and Performance

A number of variations have been systematically studied which may or may not seriously affect ratings of successful leadership in the LGD. These include: the size of the group, the seating arrangement, pretest coaching, the motivation of the participants specific to the situation, and member participation.

Effects of size. As the size of the group increases from two to twelve, the mean LGD rating assigned is reduced approximately 50 per cent. Eighty-three per cent of the variance in ratings where members come from groups of 2, 4, 6, 8, and 12 is accounted for by size according to a study of 120 examinees by the author and Norton (19). It appears that the opportunity to display successful leadership is closely associated with the size of the group. We conclude that proper correction must be made in any LGD studies where different examinees have been tested in groups varying in size.

Effects of location of seat and seating arrangement. Sixty-eight discussions among 467 participants were analyzed by the author and Klubeck (16) to determine the effects of the particular seat a participant held on the LGD rating he obtained. For both V-shaped arrangements and those in which members sat in parallel rows facing each other, members seated at the ends obtain slightly higher mean scores. In two sets of the seven studies included in this analysis, the results were significant at the 1 per cent level. The effects tended to disappear when variations in the real-life esteem of the members were held constant. At any rate, the differences, statistically significant or not, associated with participants' location in the group were too small to be of much practical concern.

Pretest coaching. If LGD performance can be greatly altered by means of brief coaching which does not reflect any real change in personality, its routine use by large organizations such as the Armed Forces for screening OCS applicants would be impossible. Therefore, Klubeck and the author (50) briefly coached the third highest and the sixth highest participants (among seven) of each of 20 leaderless discussions and then ran retests on each group of seven. An analysis of covariance led to the inference that while those who were fairly high in LGD score initially profited significantly from coaching, those who were initially low did not profit at all from the brief coaching. While the shift upward of the high ranking subjects was statistically significant, it was not very large in an absolute sense.

The investigators cited a number of reasons for rejecting the inference that the differential improvement was due to differential motivation, and concluded that LGD behavior is a function of personality traits and needs which cannot be altered readily by brief coaching.

This interpretation was in agreement with Harris' (45) opinion that one could not "cram" for the leaderless group discussion. Harris suggested that priming a candidate, rather than help would most likely handicap him by "inhibiting his spontaneity."

In an unpublished study, Pruitt and the author gave the same part-directive, part-permissive coaching

as Klubeck and the author had administered previously, but the groups receiving training were coached for 15 minutes while assembled as groups prior to undergoing the LGD. Directly in line with Harris' hypothesis, the five trained groups, each with six participants, showed significantly less "initiation" behavior than eight untrained control groups of six each. The trained groups exhibited only half as much "consideration" behavior as the untrained groups. Raters commented on the "freezing up" and the "increased nervousness and tensions" which characterized the trained groups; this was in line with Harris' hypothesis.

As pointed out by Klubeck and the author:

. . . long term training is obviously an entirely different matter. Thus, if an ineffective individual underwent psychotherapy successfully which led to favorable modifications in his needs and self-esteem, there is no reason to reject the possibility that he would exhibit improved performance on the LGD, but here the LGD would reflect real personality change (50, p. 71).

Actually, Pepinsky, Siegel, and Vanatta (61) carried out such long-term training with some success.

Effects of extrinsic motivation of participants specific to the situation. Do momentary changes in the incentive to participate, unrelated to the personality of the participants, make much difference in LGD performance? An unpublished study completed by the author tentatively suggests that added extrinsic motivation is relatively unimportant in determining the behavior of LGD participants. Two small samples of a total of 31 college students were tested and retested in leaderless groups of seven to nine. The first sample was told that the first test was mere practice and had no bearing on class grades, but that the second

test would count as an important grade determiner. The second sample took the "important" examination first and was then given the second test as a "check-up on them" that would not affect them. The means of successful leadership displayed in both the motivated and unmotivated situations were practically identical. Moreover, the correlation between LGD performance in the situations was .86, indicating that the addition of incentives affects performance relatively and absolutely very little.

Needs more basic to the personality probably energize and sustain LGD behavior. These needs appear to be either openly denied or unconscious to a great extent. When 22? ROTC cadets, in an unpublished study by the author and Coates, were asked to indicate on a five-point scale how much they tried to do as well as possible on an LGD, a correlation of only .30 was found between reported effort and actual LGD scores obtained.

Analogous to performance on paper-and-pencil aptitude tests, increasing the extrinsic motivation of mature subjects does not serve to raise scores materially, since subjects usually perform near their maximum without such added incentive. Of course, we might succeed in lowering performance if we could sufficiently discourage subjects or increase tension beyond some optimum point.

It is possible that the LGD may be no more sensitive to variations in examinee's extrinsic motivation than most aptitude tests. It is probable that the LGD is less affected by such extrinsic motivation as the desire to obtain a job than is the usual undisguised personality test.

What is needed are comparative

studies of operational, as opposed to experimental, validities of the LGD.

Amount of participation. Analyses generally find a high correlation between the sheer amount of talk of LGD participants and the scores they earn for successful leadership. Time spent talking correlated .65 with ratings of success for 64 sales and management trainee candidates (10), .77 for 140 sorority girls (23), .93 for 20 college students (6), and .96 for 36 college students of an unpublished study by R. L. French.[6]

This high correlation is disturbing at first, since it suggests that LGD ratings primarily discriminate the verbose from the terse. However, the relationship can be shown by a series of deductions to logically follow if we assume that almost all participation in the LGD is *attempted* leadership behavior, that LGD ratings are assessments of *successful* leadership behavior, and that attempted leadership must occur in order for some of it to be judged successful. More detailed discussions of these relationships are presented elsewhere (6, 13).

Kind of participation. Qualitative differences appear in the kinds of LGD participation engaged in by the successful leader and by those who participate and attempt leadership acts, but who nevertheless earn low scores as successful leaders. When the responses during leaderless discussions of 46 fraternity members were analyzed by the Bales technique (4), other judges' ratings of participants' successful leadership correlated .66 with frequency of *attempting answers;* .50 with frequency of *positive socioemotional responses;* .44

[6] FRENCH, R. L. Verbal output and leadership status in initially leaderless discussion groups. *Amer. Psychologist*, 1950, 5, 310. (Abstract)

with frequency of *asking questions;* and .32 with frequency of *negative socioemotional responses* (12). The roles associated with successful leadership were initiator-contributor, opinion-giver, elaborator, compromiser, orienter, evaluator, energizer, encourager, etc. (11). Only one type of participation, *attempting answers,* was associated (r = .48) with the real-life esteem of the participants.

In a study of 40 ROTC cadets, Carter, Haythorn, Shriver, and Lanzetta (32) found that LGD leaders were much more likely than non-leaders to diagnose the situation, ask for expressions of opinion, propose courses of action for others, support and defend their proposals, give information, express opinions, and argue with others.

RELIABILITY OF THE LEADERLESS GROUP DISCUSSION

The reliability of the LGD will be considered in two phases: rater agreement and test-retest reliability. Wherever possible we will indicate factors that systematically influence these reliability estimates.

Rater Agreement

Table 1 displays the average agreement found between any two observers in rating the first LGD administered to the designated subjects. The results suggest that the refined 7-item check list and its predecessors of 9 and 14 items used in most of the studies by the author and associates (e.g., 15) yield a consistent rater correlation of between .82 and .84. LGD scores based on two raters using the check list method yield a satisfactory estimated reliability of .90 or above.

A number of factors influence the agreement between raters. These will be considered next.

TABLE 1
Reliability of LGD Ratings Estimated by the Average Agreement between Any Two Raters

Subjects	What Rated	Rating Method	Average Correlation between Any Two Observers	Estimated Reliability of LGD Rating Using Two Observers
1. 64 sales and mgt. trainee candidates (9)	Desirability for position	Paired comparison	.67	.80
2. 219 administrative trainee candidates (3)	Personality and ability (highly intercorrelated)	Graphic rating scale	.70	.82
3. 36 male college students (30)	Leadership performance	Graphic rating scale	.61	.75
4. 40 NROTC subjects (31)	Leadership, authoritarianism, initiative, and insight (highly intercorrelated)	Graphic rating scale	.53	.70
5. 20 mixed college students (6)	Successful leadership displayed	Check list (13-item)	.67	.80
6. 46 fraternity members (20)	Successful leadership displayed	Check list (14-item)	.82	.90
7. 120 mixed college students (19)	Successful leadership displayed	Check list (9-item)	.82	.90
8. 244 ROTC cadets (15)	Successful leadership displayed	Check list (9-item)	.83	.91
9. 88 fraternity pledges (72)	Successful leadership displayed	Check list (9-item)	.84	.91
10. 140 sorority members (23)	Successful leadership displayed	Check list (7-item)	.83	.91
11. 48 mixed college students*	Initiation of structure and interaction	Check list (8-item)	.90	.95
12. 48 mixed college students*	Consideration of others	Check list (8-item)	.85	.92

* Unpublished data.

Effects of discussion effectiveness. A pronounced curvilinear relationship was found between the rated effectiveness of 67 leaderless discussions and the correlation between two observers' ratings for each discussion. Rater agreement was highest for discussions of average effectiveness and was lowest for discussions which were either extremely effective or extremely ineffective. The average eta found for four subsamples of these

67 discussions was .68 (17). It may be that observers become too interested in the very effective discussions and too bored or detached from the very ineffective ones; or, the variance of LGD ratings may be reduced in discussions at either extreme of the distribution of effectiveness, which, in turn, will serve to reduce the reliability of the ratings.

Effects of size. The number of participants in a given discussion appears to influence the extent to which observers will agree with each other. The author and Norton (19) tested five samples of 24 subjects each in discussions of two, four, six, eight, and twelve in size. Maximum agreement among observers was reached in groups of six $(r = .89)$ and was lowest in groups of two $(r = .72)$. Carter, Haythorn, Meirowitz, and Lanzetta (31) found that mean observer agreement was only .70 for four-man groups, while it was .85 when the same men were retested in eight-man groups.

Effects of test or retest. The last results cited can also be accounted for, in part, by the fact that observer agreement appears to increase when the same subjects are retested. Thus, when the 120 subjects of the author and Norton were all retested, observer agreement increased for groups of all sizes from a mean correlation of .82 to a mean correlation of .90.

Effects of the object of rating. The data summarized in Table 1 suggest that rater agreement is higher where raters employ a check list in which they merely indicate the extent to which each of a number of items of leader behavior was exhibited by each candidate, rather than where they employ a single graphic rating, or where they attempt to make inferences about the standing of the examinees on intervening variables

of personality or ability supposedly underlying the LGD behavior being observed. The median estimated reliability of check list ratings is .90; the median estimated reliability of other types of ratings is .80.

Test-Retest Reliability

The seven available test-retest reliability coefficients and the studies on which they are based are listed in Table 2 in order of the size of the coefficients. Listing them by size leads to the inference that test-retest reliability is higher the more similar the test and retest situations. The consistency of LGD behavior is higher the less group membership changes from test to retest; the less the problem changes; the less some members are increased in ability to lead; the less some members are increased in "real-life" status; the less observers are changed; and the less time between tests increases, permitting more random or biasing change to occur among participants. These results conform to the principle of consistency, proposed by the OSS Assessment Staff, that a subject will respond to similar environmental conditions in a similar manner (59).

Where changes in situation from test to retest are reduced to a minimum, a high test-retest reliability is found. It is probable that where behavior check lists describing participant behavior are used, the true test-retest reliability of the LGD is somewhere between .75 and .90.

Effects of size. The number of participants in a group appears to determine the consistency of the behavior from test to retest. For, while the author and Norton (19) found groups from four to twelve in size to have an average test-retest reliability of above .90, groups of two had a corresponding reliability of only

TABLE 2

TEST-RETEST RELIABILITY OF RATINGS OF LEADERSHIP DISPLAYED IN LEADERLESS
GROUP DISCUSSIONS RANKED IN ORDER OF SIMILARITY OF TEST AND RETEST

Subjects	Rating Method	Differences between Test Situations	Interval between Tests	Test-Retest-Correlations
1. 120 mixed college students (19)	Check list (9-item)	None	Week	.90
2. 31 mixed college students*	Check list (9-item)	One "important exam," other not	Week	.86
3. 140 sorority members (23)	Check list (7-item)	Two members of each group of 7 given training between tests	Three hours	.75
4. 25 mixed college students (24)	Participants rank each other	Groups rearranged, six retests	eek	.75
5. 20 mixed college students (6)	Checklist (13-item)	Intervening LGD among leader or followers only before second test	Six weeks	.72
6. 23 mixed college students (21)	Check list (7-item)	Group rearranged, type of problem systematically varied: case history, no problem presented, and leader specification to be outlined	Two days to week	.58
7. 172 ROTC cadets (15)	Check list (9-item)	Groups rearranged; some subjects changed in "real-life" status more than others between tests; different raters	Year	.53
8. 36 male college students (30)	Rating scale	Six intervening situational tests in groups of two before second test	Three to four months	.39

* Unpublished data.

.46. These results suggest that two-man and probably three-man leaderless group discussions should not be used for assessment purposes.

VALIDITY OF THE LEADERLESS GROUP DISCUSSION

The construct validity of the LGD will be examined in this discussion. This requires both logical and empirical review.

Figure 1 diagrams the relationships among a group of variables of importance in the study of leadership. Using a set of postulates based primarily on learning theory, the
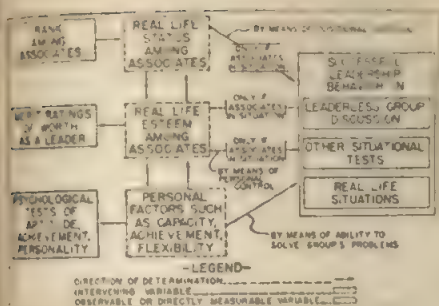
author elsewhere (13) has deduced these relationships, which may be summarized as follows:

1. The more a member *is able to solve the group's problems* because of personal characteristics (such as his capacity, achievement, responsibility, and participation), the more likely he is to exhibit successful leadership behavior in real life, in quasi-real situations, and in the leaderless group discussion. These personal characteristics are reflected in performance on various psychological tests of intelligence, proficiency, and personality.

2. The more a member exhibits successful leadership, the higher is his *esteem* among his associates—the extent to which he is regarded of worth as a member or leader to the group, regardless of his position—and the higher will be the merit ratings he receives as a successful leader or member. The higher his *esteem*, the more likely he is to be of further success as a leader among his associates.

3. The higher a member's *status*, as inferred from his rank or the worth of his *position* among his associates, the more likely he is to successfully lead his associates.

Further relations between variables noted in Fig. 1 can be ignored here.

If ratings of LGD performance are actually valid measures of tendencies of individuals to differ in successful behavior, and if we accept as a logical rule that variables with common determinants should correlate positively with each other, then following the outline of Fig. 1, LGD scores should correlate:

1. With real-life rank (and hence real-life status) when the LGD is among associates of different ranks;

2. With real-life merit ratings (hence real-life esteem);

3. With leadership performance in other quasi-real situations;

4. With observations and indices of successful leadership performance in real-life situations;

5. With personal characteristics as measured by psychological tests and measurements commonly associated with success as a leader.

Any positive correlation between an LGD rating and these other specified measures should provide partial evidence of validity of the LGD performance rating—namely that it actually measures leadership potential or individual differences in tendency to be successful as a leader. Previously presented evidence indicates that the measurement of LGD performance is consistent with itself. The question still to be answered is whether or not it is consistent with the various other measurements associated with, described as, or defined as successful leadership behavior. Rated LGD performance should be associated with these other measures if it is an assessment of success as a leader.

We will now survey empirical investigations of the extent to which rated LGD performance was found associated with status and esteem in real-life, personal characteristics and leadership performance elsewhere.

Status (as Estimated by Rank) and LGD Performance

A biserial correlation of .88 was found between the rank in the company of each of 131 oil refinery supervisors and their success as LGD leaders among their associates (22). The more the discussion problem concerned matters for which they had rank over their associates, the higher was this correlation (21). A corre-

sponding correlation of .51 was found for 264 ROTC cadets.[7] The lower correlation in ROTC probably reflected the fact that rank differences were less vital to the cadets than to the industrial executives.

When 180 ROTC cadets were retested among their associates a year after an original test, those who rose during the year from cadet noncom to cadet first lieutenant or higher gained significantly more in LGD score on the retest compared to the test than those who received promotions to cadet second lieutenant only.[8]

Esteem as Estimated by Real-Life Merit Ratings and LGD Performance

Table 3 lists 17 correlations between LGD performance and esteem-in-real-life as estimated by merit ratings. It also shows when and how the ratings of merit were obtained. The median correlation is .39 and is raised to .51 when only the seven cases in which correction was made for the unreliability of esteem ratings are considered. As shown in Table 3, LGD scores have been found moderately predictive of merit as an ROTC cadet officer (15), sorority or fraternity member (20, 23, 72), civil service administrator (3, 69), shipyard foreman (55), foreign service administrator (69), and OCS cadet officer (71). The moderate correlations between LGD scores and real-life esteem for the studies that involved discussions among strangers suggest that a common source of variance among examinees, which exists beyond the effects of situation, underlies an examinee's merit among his real-life associates and his success

[7] Bass, B. M., & Coates, C. H. Situational and personality factors in leadership in ROTC. Unpublished manuscript.

[8] See footnote 7.

as a leader among strangers. If relationship between esteem and LGD score cannot be attributed solely to the tendency of an examinee to display successful leadership among associates who esteem him. Thus the LGD appears to assess attributes of the examinee that are not specific either to the test situation or to the group in which he is tested.[9]

Variables related to the correlation between esteem and LGD performance. The total amount of successful leadership displayed in a leaderless discussion appears to reflect the average merit as leaders elsewhere of all the participants of the discussion. An analysis of 67 LGD's among fraternity pledges, sorority members, and

[9] Another interpretation of these findings has been offered by E. L. Kelly in the 1954 *Annual Review of Psychology* (Stanford, Calif.: Annual Reviews, Inc., p. 295). Kelly suggests that to some extent, the real-life merit raters may react in the same way to the same cues, irrelevant to leadership success, as do the LGD raters. Both assessments are in agreement, but the source of agreement does not necessarily concern leadership potential. Thus, in a given situation, real-life merit raters and LGD raters both may tend to assign high ratings to thin men, and logic and the literature on the subject suggest that thinness should have no relation to leadership potential.

A counterargument is as follows: The real-life merit ratings, biased as they are, are a function of the extent to which the raters value or esteem the ratees. The evaluation tends to have consequences affecting the continuing success of the performance under evaluation. If seemingly irrelevant cues such as thinness influence merit ratings in real life, they will then tend to be associated with esteem and leadership potential. In reacting to the same biasing cues, the LGD observers are in error as far as logic or psychologists are concerned, but, despite this, their error is associated with real-life leadership potential as well as with the biased real-life merit ratings.

This same counterargument will not apply where merit ratings have no future consequences on the success of the actual performance being rated, such as in the case of appraising diagnostic proficiency of physicians.

TABLE 3

CORRELATION BETWEEN LGD PERFORMANCE AND "REAL-LIFE" ESTEEM AS ESTIMATED
BY MERIT RATINGS BASED ON "REAL-LIFE" PERFORMANCE

Subjects	Tested With	Estimate of Esteem	Raters	When Esteem Data Collected in Relation to LGD	Correlation
1. 100 Army ROTC cadets (15)	Equal status associates	Rated merit as a cadet and officer	Campus cadet and tactical officer superiors	6 months before	51†
2. 96 Army ROTC cadets (15)	Equal status associates	Rated merit as a cadet and officer	Campus cadet and tactical officer superiors	6 months after	44†
3. 85 Army ROTC cadets (15)	Equal status associates	Rated merit as a cadet and officer	Summer camp tactical officer superiors	9 months after	51††
4. 55 Air ROTC cadets (15)	Equal status associates	Rated merit as a cadet and officer	Campus cadet and tactical officer superiors	6 months before	68†
5. 52 Air ROTC cadets (15)	Equal status associates	Rated merit as a cadet and officer	Campus cadet and tactical officers	6 months after	65†
6. 167 Army and Air ROTC cadets*	Equal status associates	Rated merit as a cadet and officer	Peer associates	1 year after	38
7. 46 fraternity members (20)	Equal status associates	Nominations for positions of leadership	Peer associates	1 week before	44
8. 65 fraternity pledges (72)	Strangers	Nominations for positions of leadership	Peer associates	6 months after	47†
9. 140 sorority members (23)	Strangers	Nominations for positions of leadership	Peer associates	Same time	39†
10. 168 administrator trainees (3)	Strangers	Potential capacity as administrator	Superior associates	1 year after	41
11. 84 shipyard foremen (54)	Equal status associates	Adequacy of foreman	Superior and peer associates	Not specified	29
12. 123 foreign service personnel (66)	Strangers	Suitability in foreign service	Superior associates	2 years after	41
13. 202 civil service administrators (69)	Strangers	General merit as administrator	Superior associates	2 years after	38
14. 131 oil refinery supervisors (22)	Associates of varying status	Merit as supervisor	Superior associates	1 month to 1 year before	06
15. 48 uncoached mixed college students*	Strangers of the same sex	Nominations as leader	Associates	1 and 4 weeks before	18
16. 35 coached mixed college students*	Strangers of the same sex	Nominations as leader	Associates	5-8 weeks before	43
17. 323 Air Force officer candidates (71)	Strangers of the same sex	Rated merit as a cadet and officer	OCS peers and faculty		41

* Unpublished data.

† Corrected for criterion unreliability and/or for between subsample variations in criterion means and variances.

ROTC cadets found a correlation of .35 between the mean LGD score per discussion group and the mean esteem-in-real-life score of the participants of each group. Thus, there was a between-groups as well as a within-groups positive correlation between LGD ratings and esteem as estimated by merit ratings. In the same way, the LGD scores within a discussion correlated .20 with the merit scores among the participants of that discussion (17).

These results suggest that LGD ratings which depend solely on standards within a group discussion should suffer in validity as estimated by correlations with real-life merit ratings. Rating techniques with this disadvantage include the forced distribution rating, paired comparison, ranking, and any others which force the equalization of all discussion-group LGD score means and/or variances. Similarly, when tested among strangers, any ratings of each other by the participants themselves will be attenuated in validity as predictions of esteem, since they will depend solely upon standards based on observation of a single discussion.

This same analysis (17) indicated that a number of variables are associated with the variation from discussion to discussion in the correlation between real-life esteem of the members and their LGD performance. According to a Doolittle solution, these variables included: within-discussion variance in real-life esteem; within-discussion variance in LGD ratings; and group size. All relationships were positive except for size. Six-man groups were slightly more valid than larger ones as predictors of real-life esteem.

Status differences among members almost completely invalidate the LGD as an indicator of real-life esteem. Yet above and beyond these

effects, the case history discussion appears more likely than other types of discussions to reflect differences in real-life esteem and personality. Where members of different rank were tested together, the case history discussion was the only one which yielded scores that correlated positively with merit ratings as refinery supervisors ($r = .28$). Furthermore, case history discussion performance correlated .54 with a supervisory aptitude test battery. Other types of discussion, in these circumstances, averaged .23 in correlation with supervisory aptitude test scores, and correlated negatively with ratings of esteem of these supervisors of different rank (21).

Personal Characteristics Associated with Leadership and LGD Performance

According to Stogdill's survey of over one hundred studies, leaders tend to surpass nonleaders in certain personal characteristics such as *capacity* (intelligence, alertness, verbal facility, originality, and judgment), *achievement* (scholarship and knowledge), *responsibility and associated personality factors* (dependability, initiative, persistence, aggressiveness, self-confidence, desire to excel), and *participation* (activity, sociability, cooperation, and adaptability) (63). If these characteristics can be included under the broad concept of "abilities to solve group problems," then their relationship to leadership can be deduced as well (13). Performance in the LGD should be associated with these personal factors, if it is to be judged valid as a measure of individual differences in tendency to exhibit successful leadership behavior. Table 4 shows the correlations obtained between various measures of capacity and/or achievement and LGD performance. Table 5

TABLE 4
CORRELATION BETWEEN LGD PERFORMANCE AND MEASURES OF CAPACITY AND ACHIEVEMENT

Subjects	Test of Capacity or Achievement	Factor(s) Most Probably Measured by Test or Measurement	Correlation
1. 140 sorority members (23)	ACE Linguistic	Verbal aptitude	.35
2. 66 fraternity pledges (72)	ACE Linguistic	Verbal aptitude	.32
3. 64 sales and management trainee candidates (10)	OSPE	Verbal aptitude	.25
4. 140 sorority members (23)	ACE Quantitative	Numerical aptitude	.17
5. 66 fraternity pledges (72)	ACE Quantitative	Numerical aptitude	.17
6. 180 ROTC cadets*	ACE Total	Verbal and numerical aptitude (intelligence)	.25
7. 202 administrator candidates (68)	South African Air Force test of mental alertness	Intelligence	.32†
8. 131 oil refinery supervisors (22)	Otis Gamma	Verbal, spatial, numerical, aptitude (intelligence)	.45
9. 123 foreign service candidates (69)	Cognitive Test Battery	Intelligence	.32
10. 202 administrator candidates (69)	Cognitive Test Battery	Intelligence	.20
11. 123 foreign service candidates (69)	Civil Service Qualifying Exam	Intelligence and scholastic achievement	.30
12. 202 administrator candidates (69)	Civil Service Qualifying Exam	Intelligence and scholastic achievement	.20
13. 140 sorority members (23)	Years of education	Scholastic achievement	.18
14. 131 oil refinery supervisors (22)	Years of education	Intelligence and scholastic achievement	.57†
15. 180 ROTC cadets	Average grade in college	Scholastic achievement	.16
16. 140 sorority members (23)	Average grade in college	Scholastic achievement	.31
17. 131 oil refinery supervisors (22)	Supervisory aptitude test	Supervisory aptitude	.30‡
18. 66 mixed college students (64)	How Supervise?	Supervisory knowledge	.46

* Unpublished data.
† Contaminated by the correlation between rank and test of measurement.
‡ Not affected by holding rank constant.

shows the correlations between LGD performance and various personality variables that approximate the "responsibility" and "participation" clusters of Stogdill. The first three items of Table 7 may be regarded as further evidence of the correlation between participation and LGD performance. We shall briefly consider, in turn, the correlations between LGD performance and capacity, achievement, responsibility, and participation.

Capacity and leaderless group discussion performance. While the correlation between LGD performance and verbal aptitude averages .30 or above, as might be expected, the correlation between LGD behavior and numerical aptitude is below .20.

Intelligence test scores measuring verbal, numerical, and spatial factors tend to fall in between in correlation with LGD performance. Ability to solve the leaderless discussion group's problems appears to depend more on verbal than on other aptitude factors.

A factor analysis of 14 leadership and ability measures by the author and Coates[10] found that while rated performance in initially leaderless discussions correlated close to zero with one factor, Ability in Active Situations, it correlated .44 with another factor, Ability in Verbal Situations. Sakoda's (62) analysis of OSS data also noted that discussion ratings fell into a cluster with other ratings based on "verbal" situations. Carter, Haythorn, and Howell (30) arrived at the same conclusion following factorial analyses of several criteria of leadership. These findings may indicate the boundaries to the range of real-life situations in which leadership behavior can be forecast by rated LGD performance.

The 10 available correlations between intelligence and associated aptitude test scores and success in the LGD appear consistent with expectations. Yet, verbal aptitude only accounts for 10 to 15 per cent of the variance in LGD, and so cannot be regarded as a more easily administered substitute for predicting success as a leader.

Achievement and leaderless group discussion performance. Six studies are available of the correlation between LGD ratings and tested achievement, years of education, or grades in college. Again, the correlations are uniformly positive, ranging from .16 to .31, with a median of .25. If the unusually high correlation of .57 is ignored because it is contaminated by the correlation of rank with both education and LGD success, the median becomes .20.

Above and beyond the effects of rank, a correlation of .30 appears to exist between LGD performance and supervisory aptitude as measured by an optimally weighted battery of interest, biographical, and supervisory judgment tests.

How Supervise?, a fairly widely used test of knowledge of principles of supervision, which aims to predict success as an industrial supervisor, correlates .46 with the LGD success of 66 college students (64).

Personality and leaderless group discussion performance. We shall now consider correlations obtained between LGD performance and the specific personality traits found to be associated with leadership according to Stogdill.

Stogdill cites five studies in which leaders are found to be more energetic than nonleaders. For the LGD, a correlation of .15 was found between general activity and energy as assessed by the Guilford-Zimmerman Temperament Survey for 76 sorority girls. A corresponding correlation of .12 was found for 66 college students (64). Also, LGD leaders were more often characterized in a Rorschach analysis as highly energetic, while LGD nonleaders more often were described as lazy or passive (18).

Stogdill uncovered a large number of studies that found successful leadership associated with originality, soundness of judgment, and ability to evaluate situations. Rorschach analysis characterized LGD leaders as strongly imaginative, strongly interested in details, and able to see the larger aspects of things, and LGD nonleaders as stereotyped or conventional in thoughts and perceptions, as unclear, plodding, and confused thinkers, and as unimaginative. A

10 See footnote 7.

study of 172 ROTC cadets found an eta of .30 between the F scale of authoritarianism and LGD performance. Highly authoritarian—hence stereotyped and rigid—personalities did extremely poorly in the discussion, while equalitarian—but not too equalitarian—cadets earned the highest LGD scores. Among samples of 100 and 67 of these same cadets, Pearson correlations of .32 and .33 were found between LGD performance and Thurstone's Concealed Figures and Gestalt Completion tests —tests of perceptual flexibility.[11]

Self-assurance and absence of modesty were uniformly associated with leadership in 17 studies cited by Stogdill. Similarly, self-esteem as measured by 140 self-nominations for sorority leadership positions correlated .29 with LGD performance. An analysis of interviews with nine LGD leaders and nine LGD nonleaders from a total of 140 subjects using the Who-Are-You technique (28) showed that compared with nonleaders, LGD leaders more frequently regard in a more favorable light themselves, their effects on others, and other persons' effects on them.

Discrepancies between Stogdill's conclusions concerning leadership in general and LGD personality correlations arise when we consider personality characteristics such as emotional stability, sociability, ascendency, and responsibility.

Eleven studies have found emotional stability to be associated with leadership, although the evidence is not uniformly positive (63). Similarly, LGD performance correlated .20 and .17 with emotional stability as measured by the Guilford-Zimmerman inventory (18, 64). Rorschach analysis likewise characterized more LGD leaders than nonleaders as

emotionally stable. Overt social adjustment based on peer ratings correlated .28 with LGD performance. On the other hand, five correlations found between LGD performance and the inventoried traits of freedom from hypersensitivity and hostility ranged from −.40 to .29 with a median of −.04 (18. 64).

Responsibility was uniformly reported by Stogdill to be associated with leadership. But a correlation of −.29 was found for 47 ROTC cadets between LGD ratings and responsibility as measured by the Gordon Personal Profile, while correlations of .19 and .26 were found in two analyses of the relations between thoughtfulness as assessed by the Guilford-Zimmerman and LGD performance (18, 64).

Evidence concerning the relation between extroversion, ascendency, and leadership in general is contradictory according to Stogdill (63). Ascendency as measured by the Guilford-Zimmerman Temperament Survey correlated .44 with LGD performance of 76 sorority girls (18) and .25 with LGD performance of mixed college students (64); but ascendency as measured by the Gordon Personal Profile and the A-S Reaction Test correlated only .02 and −.02, respectively, with LGD performance.

Sociability as measured by the Guilford-Zimmerman in two studies (18, 64) correlated .27 and .31, respectively, with LGD performance; similar correlations between cooperativeness and LGD success were .14 and .13, while sociability as measured by the Gordon Personal Profile correlated close to zero with LGD performance.

Finally, while Stogdill found ambition and desire to excel important attributes of leadership, an average correlation of −.05 was found between LGD performance and ex-

[11] See footnote 7.

TABLE 5

CORRELATION BETWEEN LGD PERFORMANCE AND PERSONALITY TESTS OR MEASUREMENTS

Subjects		Personality Test or Measurement	Trait Probably Measured	Correlation
1.	76 sorority members (18)	Guilford-Zimmerman Temperament Survey—G	General activity and energy	.15
2.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—G	General activity and energy	.12
3.	20 sorority members highest and lowest on LGD (18)	Rorschach	Imagination, strongly interested in details, able to see larger aspects of things vs. conventionality, stereotypy, etc.	.60
4.	172 ROTC cadets (15)	UCPOC F scale	Authoritarianism, rigidity	.30†
5.	100 ROTC cadets*	Concealed Figures	Perceptual flexibility	.32
6.	67 ROTC cadets*	Gestalt Completion	Perceptual flexibility	.33
7.	140 sorority members (23)	Self-nominations for leadership positions in sorority	Self-esteem	.29
8.	18 sorority members highest and lowest on LGD (18)	W-A-Y interview	Self-esteem	‡
9.	76 sorority members (18)	Guilford-Zimmerman Temperament Survey—E	Emotional stability	.20
10.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—E	Emotional stability	.17
11.	140 sorority members (23)	Peer ratings	Overt social adjustment	.28
12.	49 sorority members (18)	Guilford-Zimmerman Temperament Survey—F	Friendliness, agreeableness	−.23
13.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—F	Friendliness, agreeableness	.04
14.	47 ROTC cadets*	Gordon Personal Profile—H	Freedom from hypersensitivity	−.40
15.	49 sorority members (18)	Guilford-Zimmerman Temperament Survey—O	Freedom from hypersensitivity	−.04
16.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey · O	Freedom from hypersensitivity	.29
17.	49 sorority members (18)	Guilford-Zimmerman Temperament Survey—P	Cooperativeness	.14
18.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—P	Cooperativeness	.13
19.	76 sorority members (18)	Guilford-Zimmerman Temperament Survey—S	Sociability	.27

TABLE 5—Continued

	Subjects	Personality Test or Measurement	Trait Probably Measured	Corre-lation
20.	66 Mixed college students (64)	Guilford-Zimmerman Temperament Survey—S	Sociability	.31
21.	47 ROTC cadets*	Gordon Personal Profile—S	Sociability	.07
22.	47 ROTC cadets*	Gordon Personal Profile—R	Responsibility	−.29
23.	49 sorority members (18)	Guilford-Zimmerman Temperament Survey—T	Thoughtfulness, reflectiveness	.19
24.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—T	Thoughtfulness, reflectiveness	.26
25.	76 sorority members (18)	Guilford-Zimmerman Temperament Survey—A	Ascendency	.44
26.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—A	Ascendency	.25
27.	47 ROTC cadets*	Gordon Personal Profile—A	Ascendency	.02
28.	66 mixed college students (64)	A-S Reaction Test	Ascendency	−.02
29.	140 sorority members (23)	Ratings of motivation for various offices	Motivation to lead	−.05
30.	46 uncoached college students*	Check list descriptions	Parental initiation	−.08
31.	30 coached college students*	Check list descriptions	Parental initiation	.19
32.	46 uncoached college students*	Check list descriptions	Parental consideration	−.07
33.	30 coached college students*	Check list descriptions	Parental consideration	.07
34.	39 uncoached college students*	Kerr Empathy Test	Social knowledge	.16
35.	29 coached college students*	Kerr Empathy Test	Social knowledge	−.19
36.	22 coached college students*	Dymond Empathy Test (modified)	Accuracy of estimated self-ratings of others	−.07
37.	22 coached college students*	Dymond Empathy Test (modified)	Accuracy of estimated group ratings of self	.36
38.	22 coached college students*	Dymond Empathy Test (modified)	Accuracy of estimated group ratings of others	.19
39.	49 sorority members (18)	Guilford-Zimmerman Temperament Survey—M	Masculinity-femininity	−.02
40.	66 mixed college students (64)	Guilford-Zimmerman Temperament Survey—M	Masculinity-femininity	−.01

* Unpublished data.
† Curvilinear relationship. Maximum LGD success experienced by equalitarian—but not too equalitarian—subjects. Highly authoritarian subjects perform worst on LGD.
‡ Chi-square analysis suggested that a significant positive correlation existed between the self-esteem of interviewees and their LGD scores.

pressed desire to hold sorority or university student offices (23).

It may be inferred that some consistency exists between Stogdill's generalizations of the relations between leadership in general and such personality traits as energy, flexibility of judgment, and self-esteem, and the relations between these traits and LGD performance. Contradiction or lack of uniformity appears when we consider such traits as responsibility, emotional stability, ascendency, and sociability.[12]

Participation and LGD performance. According to Stogdill, leaders,

[12] Part of this lack of uniformity may be due to variations among techniques used to measure the various personality traits, and variations in the sex composition of the samples studied.

The responsibility, ascendency, sociability, and freedom from hypersensitivity scales of the Gordon Personal Profile, a forced-choice personality inventory, administered to samples of men only, tended to correlate zero or negatively with LGD success. Corresponding Guilford-Zimmerman scales with the same or similar names—of the traditional self-report type—tended to correlate positively with the LGD success of a sample of women only (18) and with the LGD success of mixed men and women (64).

Assuming that Gordon's forced-choice procedure is less subject to distortion than the Guilford-Zimmerman, and assuming that the measurement techniques rather than the samples were a significant source of variance, one might speculate that LGD performance tends to be associated with a participant's concept of himself, but only where the participant is free to distort the description to suit himself.

To go one step further, it may be that LGD success is related to the way a participant likes to see himself, but not to the way he actually sees himself when forced to make discriminations over which he has less control.

Neither of these self-inventoried evaluations actually meets the requirements of the original hypothesis that the more able member is more likely to be esteemed and to be successful as a leader. The crucial evaluations of ability for leadership and esteem are those based not on self-evaluation, but on other group members' judgments of the participant, or on objective *tests* of personality (as contrasted with inventories).

in general, tend to be more talkative, more industrious, and more likely to participate in group activities. The same appears to be true of high scorers on the LGD who, compared with low scorers, were found to make 1.5 times as many responses in a personal interview, and to give 1.6 as many responses to the Rorschach (18). LGD success also has been found, for 140 sorority members, to correlate .36 with the number of university student leadership positions held per semester, and .10 with the number of sorority positions held per semester (23).

Parental leadership and LGD performance. It is expected that childhood experiences and memories of them should play a significant role in adolescent or adult leadership behavior (13). In an unpublished study, the author hypothesized that LGD performance would be a function of the participants' perception of how they had been led by their parents. Examinees used modified Ohio State Leadership Studies behavior check lists (38) to describe the extent to which their mothers and fathers initiated structure for them and were considerate of them. The average and range of correlations between such descriptions and LGD performance suggested that no consistent relationship existed between parental descriptions and LGD performance.

Empathy and LGD performance. A series of studies (e.g., 33, 67) has found a moderate relationship between empathic ability and leadership. The results have not been uniformly positive, mainly because the measure of empathy has varied greatly from one investigation to another. Elsewhere (13), the author has deduced that the more a person can accurately estimate the *needs* of others, the more likely he is to successfully lead others. Chowdry and

Newcomb (33) have come closest to testing this hypothesis, and have obtained positive results.

Kerr and Dymond Empathy test data collected on a small sample (N's = 22 to 39) of students by Stolper (64) were correlated by the author with LGD performance scores, as shown in Table 5. The correlation of .36 between *accuracy of estimations of group ratings of self* on the Dymond test and LGD performance was somewhat artifactual since LGD leaders probably led in the Dymond test also, and *all* members of a group are likely to agree more closely on leaders' ratings compared to non-leaders, according to an earlier study by the author (6).

Leadership Performance in Other Quasi-Real Situations and LGD Performance

According to Fig. 1, leadership in other quasi-real situations is governed by the same individual factors as performance in the LGD. Therefore, the two should correlate fairly highly with each other if successful leadership is being measured in these other situations, and if ratings of LGD performance are valid as leadership ratings. Similar deductions can be made about LGD performance and success as a leader in real life.[13]

Table 6 summarizes the 17 available correlations between leadership or "desirability for leadership positions" ratings based on LGD's and other situational tests. The correlations are almost uniformly highly positive, but many are contaminated because the same raters were used to assess candidates during the LGD and the other situations.

The median correlation between interview and LGD ratings is .70.

[13] The same alternative interpretation and rebuttal apply here as are presented in footnote 6.

Where ratings are made by different raters in each situation, the correlation drops to .45 (9).

Correlations of .37 and .67 were found between ratings based on assigned leadership and leaderless discussion (3, 59). Correlations between the LGD and other verbal situational tests, such as debates and committee work, are almost as high as between the LGD and its retest, averaging .70. As the other situations involve less discussion and more mechanical or athletic activity, the median correlation between the LGD and these other situational tests is reduced to around .52.

Leadership ratings based on the leaderless group discussion alone correlated .64 with the final leadership ratings based on the entire OSS battery of situational tests (59).

Leadership Performance in Real-Life Situations and LGD Performance

Table 7 summarizes the correlations between LGD performance and real-life leadership performance.[14] While LGD ratings correlate somewhat with the tendency to hold leadership offices, they appear to be associated more with the tendency in real life to initiate structure and interaction among associates and subordinates ($r = .32$). On the contrary, a low negative correlation exists between LGD performance and the tendency in real life to be considerate

[14] An attempt has been made in this article to treat separately from the more numerous studies of LGD performance correlated with merit ratings of real-life performance (Table 3), studies of the LGD correlated with objective indices of real-life leadership performance or fairly nonevaluative descriptions of real-life leadership performance. While usually empirically related, merit rating of performance is considered conceptually independent of actual performance or descriptions of performance.

TABLE 6

CORRELATION BETWEEN PERFORMANCE IN OTHER QUASI-REAL SITUATIONS AND IN THE LGD

Subjects		Other Quasi-Real Situational Tests	Trait Measured	Correlation
1. 223–442	OSS applicants (59)	Interview	Leadership	.48
2. 123	foreign service candidates (69)	Interview	Desirability for job	.75
3. 202	administrator candidates (69)	Interview	Desirability for job	.78
4. 64	sales management trainees (9)	Interview with different raters	Desirability for job	.45
5. 14	shoe factory trainee executive candidates (65)	Interview and tests	Desirability for job	.70
6. 223–442	OSS applicants (59)	Assigned leadership	Leadership	.37
7. 219	administrator candidates (3)	Assigned leadership	"Personality and ability"	.67
8. 223–442	OSS applicants (59)	Debate	Leadership	.56
9. 223–442	OSS applicants (59)	Brook	Leadership	.47
10. 223–442	OSS applicants (59)	Construction	Leadership	.30
11. 40	NROTC subjects (31)	Leaderless Mechanical Assembly—4-man	Leadership	.37
		Leaderless Mechanical Assembly—8-man	Leadership	.58
		Leaderless Group Reasoning Task—4-man	Leadership	.62
		Leaderless Group Reasoning Task—8-man	Leadership	.78
12. 123	foreign service candidates (69)	Other situational tests such as committee work	Desirability for job	.75
13. 202	administrator candidates (69)	Other situational tests such as committee work	Desirability for job	.70

of the welfare of subordinates and associates ($r = -.25$).[15]

Studies of Factors Associated with LGD Performance

Two factorial studies analyzing LGD test and retest scores in a correlation matrix of real-life, situational test, and psychological test measures are available. The first (23) was based on 41 measurements of 140 sorority girls; the second concerned 14 measures made of 66 to 244 ROTC cadets.[16] Table 8 lists the loadings of the LGD test and retest on the factors that accounted for most of the variance of the LGD.

In line with propositions stated earlier, ratings based on LGD performance appear to assess the extent to which an individual initiates structure or is socially bold in ambiguous situations. Esteem, or personal worth, and verbal ability are also involved. On the basis of these studies it may be inferred that there

[15] See footnote 7.

[16] See footnote 7.

TABLE 7

CORRELATION BETWEEN LGD PERFORMANCE AND REAL LIFE BEHAVIOR

Subjects	Method of Measurement	Real-Life Behavior Measured	Correlation
1. 140 sorority girls (23)	Extent of extracurricular activities	Participation in social groups	.43
2. 140 sorority girls (23)	Number of university leadership positions held	Real-life leadership performance	.36
3. 140 sorority girls (23)	Number of sorority leadership positions held	Real-life leadership performance	.10
4. 140 sorority girls (23)	Extent peers know her well enough to rate	Visibility among associates	.24
5. 180 ROTC cadets*	Final cadet rank achieved	Future success as a leader	.27†
6. 133 ROTC cadet officers*	Ratings of subordinates and peer associates	Degree of "consideration of subordinates"	−.25
7. 133 ROTC cadet officers*	Ratings of subordinates and peer associates	Degree of initiation of structure and interaction among associates and subordinates	.32
8. 32 sales trainee candidates (9)	Number of leadership positions held previously	Real-life leadership performance	‡

* Unpublished data.
† Eta (curvilinear relationship).
‡ Significant difference at 15 per cent level between LGD score of experienced leaders and those without experience as leaders.

exist three independent sources of variance underlying leaderless group discussion performance:

I. Tendency to Initiate Structure

II. Tendency to Be Esteemed or of Value to a Group

III. Ability in Verbal Situations.

Little specific variance is left when the common variance accounted for by these factors is extracted.

SUMMARY

The history, applicability, reliability, and validity of the leaderless group discussion as a means of assessing variations among persons in the tendency to exhibit successful leadership behavior have been considered. While the procedure was originated as a psychological technique in Germany over thirty years ago, it is

TABLE 8

ORTHOGONAL FACTORS CORRELATED WITH OVER-ALL LGD PERFORMANCE

Factor	Correlation in Sororities (23)	Factor	Correlation in ROTC*
I. Leadership Potential (Esteem)	.24	I. Esteem	.30
II. Ascendency-Sociability	.39	II. Tendency to Initiate Structure	.53
III. Verbality	.51	III. Ability in Verbal Situations	.44
IV. Intellectualism	.21		

* Unpublished data.

only in the last decade that systematic reliability and validity studies have appeared.

High interrater agreement and high test-retest reliabilities have been reported consistently, especially where descriptive behavior check lists have been used as the rating technique.

Group size, length of testing time, type of problem presented, directions, seating arrangement, number of raters, and rating procedure influence to a greater or lesser extent performance in the LGD as well as the reliability and validity of LGD ratings. Studies of the effects of many of these have appeared since 1950.

According to both deductive and inductive evidence, a valid assessment of the tendency to display successful leadership should correlate with: (a) status as measured by rank, when the assessment is based on performance among associates of different rank; (b) esteem in real life as estimated by merit ratings; (c) successful leadership performance in other quasi-real and real-life situations; (d) personal characteristics as measured by psychological tests, such as capacity, proficiency, responsibility, and participation. On the whole, ratings based on performance in the LGD tend to do this. Therefore, it is inferred that they have some validity as assessments of the tendency to display successful leadership, i.e., leadership potential.

Other evidence suggests that the successful leadership behavior observed in the LGD concerns primarily initiation of structure rather than consideration of the welfare of others.

The preceding analysis, coupled with recommendations made by others working in the field of situational tests, such as Weislogel (71), leads us to the following hypotheses:

1. To maximize the reliability and validity of the LGD and other situational tests, scoring techniques should *minimize* reliance on the ability of observers to infer differences in personality traits and future tendencies among examinees. Observers should merely report or evaluate the immediate behavior they observe. For example, in an unpublished study, the author found that two Army colonels' estimates of the potential as Army officers of ROTC examinees were less valid as predictors of the merit ratings of the examinees than were the colonels' check list descriptions of who initiated structure during the LGD. Similar results were noted in a study of fraternity members reported by the author and White (20).

When the observer makes an inference about the future behavior of an examinee from the observations of the examinee during the LGD, several potential errors are likely. The observer may err in deciding on which dimensions to make inferences; he may err in collating his observations with the future behavior to be predicted; and finally, the dimensions on which the inferences are made may be private ones which cannot be shared with other observers. The errors may be constant, variable, or both. Lack of knowledge and control over such errors disappears when raters are merely asked to describe what they observed and these descriptions are used as predictors.

Further reduction of uncontrolled raters' errors may be made in the following ways:

a. Objective criteria for describing specific behaviors can be used (71). In the LGD, the actual number of times a participant suggests a new approach to a problem can be noted

instead of rating "to what extent did the participant suggest new approaches to problems."

b. Forced-choice check lists can be used instead of present check lists. Otis' (60) recent successful application of the forced-choice technique to interviewer ratings indicates promise for applying the same procedure to the LGD.

2. To maximize validity, problems that are equally ambiguous to all participants, and that require the initiation of structure for their solution, should be used. Where interest is in forecasting leader behavior in real life, the structure to be set up should approximate the real-life setting as much as possible.

3. Since the LGD correlates fairly highly with most other intellectual or verbal situational tests, the use of many situational tests in a battery to forecast leadership potential is of doubtful utility. Thus, leadership ratings based on a one-hour LGD correlated above .60 with leadership assessments based on the three days of OSS situational testing (59). A similar correlation between the LGD and an entire battery of situational tests was found by Vernon (69). However, a significant proportion of the variance in over-all potential as a successful leader, unaccounted for by LGD, may be predicted by a fairly pure active or mechanical, initially leaderless, situational test which minimizes variance due to verbal ability.

4. Compared to paper-and-pencil techniques, the LGD is expensive; compared to the individual interview, in many locales, it may prove economical. The LGD appears feasible administratively, especially in military programs screening OCS or advanced ROTC applicants, in civil service examinations, in screening college seniors who are to be assessed at their colleges for management trainee positions, and anywhere else where "boards" have been used traditionally, such as in the selection of public school teachers.

5. While the LGD appears to have some validity as a predictor of the tendency to be a successful leader in a number of situations, especially in comparison to other assessment techniques, tailor-made batteries of paper-and-pencil tests will undoubtedly yield higher validities in designated situations. However, it may be that just as the brief intelligence test is applicable for predicting trainability for many skilled occupations, so the LGD will provide a general technique for partially assessing potential success as a leader in a relatively wide range of situations.

6. A number of situations in which the LGD is less likely to be successful may include the following:

a. The LGD is less likely to be valid for measuring or forecasting esteem or leadership potential when examinees can be tested only among others of different rank. In such a case, status—and not esteem or personality—will determine who succeeds in the LGD.

b. The LGD is less likely to be useful where factors peculiar to the situation block initiation of structure where no structure exists. Conceivably, in certain military settings for example, the examinees may be imbued with the dictum "never volunteer for anything." However, exactly how this would affect LGD validities is unknown.

c. Another unknown is the effect of the average verbal aptitude and educational status of the participants on the validity and utility of the LGD. It is expected that where this mean falls below a certain minimum,

LGD forecasting efficiency may suffer.[17]

d. Since achievement and intelligence appear to correlate with LGD performance as well as with success as a leader in real life, any restriction in the range of intelligence or achievement of LGD participants would be likely to reduce the forecasting efficiency of the LGD. Conversely, the greater the variance in intelligence and achievement of participants, the more likely is the LGD to accurately assess leadership potential.

e. Finally, the less the leadership situation for which we are selecting leaders requires verbal communication or verbal problem solving, the less likely is the LGD to be useful as a measure or predictor of the tendency to exhibit successful leadership.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. Ansbacher, H. L. The history of the leaderless group discussion technique. *Psychol. Bull.,* 1951, **48**, 383–390.
3. Arbous, A. G., & Maree, J. Contribution of two group discussion techniques to a validated test battery. *Occup. Psychol.,* 1951, **25**, 1–17.
4. Bales, R. *Interaction process analysis.* Cambridge: Addison Press, 1950.
5. Barnard, M. A., & Brody, W. A. A new method of selecting health officers-in-training. *Amer. J. Publ. Hlth,* 1946, **37**, 716.
6. Bass, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.,* 1949, **33**, 527–533.
7. Bass, B. M. Selecting personnel by observation. *Personnel,* 1950, **26**, 269–272.
8. Bass, B. M. The leaderless group discussion technique. *Personnel Psychol.,* 1950, **3**, 17–32.
9. Bass, B. M. Situational tests: I. Individual interviews compared with leaderless group discussions. *Educ. psychol. Measmt,* 1951, **11**, 67–75.
10. Bass, B. M. Situational tests: II. Variables of the leaderless group discussion. *Educ. psychol. Measmt,* 1951, **11**, 196–207.
11. Bass, B. M. Discussion and external leadership status related to roles played in initially leaderless group discussions. *Amer. Psychologist,* 1951, **6**, 486. (Abstract)
12. Bass, B. M. Differential response patterns in initially leaderless discussions related to discussion and external status. *Amer. Psychologist,* 1951, **6**, 511. (Abstract)
13. Bass, B. M. A psychological theory of leadership. Mimeographed manuscript, Louisiana State Univer., Baton Rouge, 1953.
14. Bass, B. M. Symposium on situational performance tests: the leaderless group discussion as a leadership evaluation technique. *Personnel Psychol.,* in press.
15. Bass, B. M., & Coates, C. H. Forecasting officer potential using the leaderless group discussion. *J. abnorm. soc. Psychol.,* 1952, **47**, 321–325.
16. Bass, B. M., & Klubeck, S. Effects of seating arrangement on leaderless group discussions. *J. abnorm. soc. Psychol.,* 1952, **47**, 724–727.
17. Bass, B. M., Klubeck, S., & Wurster, C. R. Factors influencing the reliability and validity of leaderless group discussion assessment. *J. appl. Psychol.,* 1953, **37**, 26–30.
18. Bass, B. M., McGehee, C. R., Hawkins, W. C., Young, P. C., & Gebel, A. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.,* 1953, **48**, 120–128.
19. Bass, B. M., & Norton, F. T. M. Group size and leaderless discussions. *J. appl. Psychol.,* 1951, **6**, 397–400.
20. Bass, B. M., & White, O. Situational tests: III. Observers' ratings of leaderless group discussion participants as indicators of external leadership status. *Educ. psychol. Measmt,* 1951, **11**, 355–361.

[17] Subsequent to the preparation of this manuscript, the author found no systematic correlation between mean ACE scores and the validity (as estimated by correlation with real-life esteem) of 64 leaderless group discussions among college students.

21. Bass, B. M., & Wurster, C. R. Effects of the nature of the problem on LGD performance. *J. appl. Psychol.*, 1953, **37**, 96–99.

22. Bass, B. M., & Wurster, C. R. Effects of company rank on LGD performance of oil refinery supervisors. *J. appl. Psychol.*, 1953, **37**, 100–104.

23. Bass, B. M., Wurster, C. R., Doll, P. A., & Clair, D. J. Situational and personality factors in leadership among sorority women. *Psychol. Monogr.*, 1953, **67**, No. 16 (Whole No. 366).

24. Bell, G. B., & French, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.

25. Benne, K. D., & Sheats, P. Functional roles of group members. *J. soc. Issues*, 1948, **4(2)**, 41–49.

26. Brody, W. Judging candidates by observing them in unsupervised group discussion. *Personnel*, 1947, **26**, 170–173.

27. Brody, W., & Powell, N. J. A new approach to oral testing. *Educ. psychol. Measmt*, 1947, **7**, 289–298.

28. Bugental, J. F. T., & Zelen, S. L. Investigations into the 'self-concept.' I. The W-A-Y technique (Who are you?). *J. Pers.*, 1950, **18**, 483–498.

29. Carter, L. F. Recording and evaluating the performance of individuals as members of small groups. *Amer. Psychologist*, 1953, **8**, 464. (Abstract)

30. Carter, L., Haythorn, W., & Howell, Margaret. A further investigation of the criteria of leadership. *J. abnorm. soc. Psychol.*, 1950, **45**, 350–356.

31. Carter, L., Haythorn, W., Shriver, Beatrice, & Lanzetta, J. The relation of categorizations and ratings in the observation of group behavior. *Hum. Relat.*, 1951, **4**, 239–253.

32. Carter, L., Haythorn, W., Shriver, Beatrice, & Lanzetta, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

33. Chowdry, K., & Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.

34. Couch, A. S., & Carter, L. F. A factorial study of the rated behavior of group members. *Amer. Psychologist*, 1952, **7**, 537. (Abstract)

35. Douglas, A. G. Shall civil service endorse science or novelty? *Publ. Adm. Rev.*, 1950, **10**, 78–86.

36. Fields, H. The group interview test: its strength. *Publ. Personnel Rev.*, 1950, **11**, 139–146.

37. Fields, H. An analysis of the use of the group oral interview. *Personnel*, 1951, **27**, 480–486.

38. Fleishman, E. A. The description of supervisory behavior. *J. appl. Psychol.*, 1953, **37**, 1–6.

39. Fraser, J. M. An experiment with group methods in the selection of trainees for senior management positions. *Occup. Psychol.*, 1946, **20**, 63–67.

40. Fraser, J. M. New-type selection boards in industry. *Occup. Psychol.*, 1947, **21**, 170–178.

41. Garforth, G. I. de la P. War officer selection boards. *Occup. Psychol.*, 1945, **19**, 97–108.

42. Gellhorn, W., & Brody, W. Selecting supervisory mediators through trial by combat. *Publ. Adm. Rev.*, 1948, **8**, 259–266.

43. Gibb, C. A. The principles and traits of leadership. *J. abnorm. soc. Psychol.*, 1947, **42**, 267–284.

44. Gulliksen, H. Intrinsic validity. *Amer. Psychologist*, 1950, **5**, 511–517.

45. Harris, H. *The group approach to leadership-testing.* London: Routledge & Kegan-Paul, 1950.

46. Hemphill, J. K. Leadership in small groups. Multilithed manuscript, Personnel Research Board, The Ohio State Univer., 1952.

47. Herbert, L. A. Case study of a group selection procedure for selecting candidates for teacher's training. *Hum. Relat.*, 1951, **4**, 77–94.

48. Hollander, E. P. Authoritarianism and leadership choice in a military setting. *Amer. Psychologist*, 1953, **8**, 368. (Abstract)

49. Kelley, H. H. Communication in experimentally created hierarchies. *Hum. Relat.*, 1951, **4**, 39–56.

50. Klubeck, S., & Bass, B. M. Differential effects of training on persons of different leadership status. *Hum. Relat.*, 1954, **7**, 59–72.

51. Landry, H. A., Krugman, M., & Wrightstone, J. W. *Validation study of the group oral interview test.* New York: Bd. of Educ., 1951.

52. Lang, H. A. A statement on the practical application of role playing in the group interview technique by the Washington

State Personnel Board. *Sociometry*, 1951, 14, 63-68.

53. MANDELL, M. M. The group oral performance test. *Publ. Personnel Rev.*, 1946, 4, 209-212.

54. MANDELL, M. M. Testing for administrative and supervisory positions. *Publ. Personnel Rev.*, 1948, 9, 190-193.

55. MANDELL, M. M. Validation of group oral performance test. *Personnel Psychol.*, 1950, 3, 179-185.

56. MANDELL, M. M. The group oral performance test. Multilithed manuscript, U. S. Civil Service Commission, 1952.

57. MANDELL, M. M. The group oral performance test. *Personnel Adm.*, 1953, 15, 11-17.

58. MEYER, C. A. The group interview test: its weakness. *Publ. Personnel Rev.*, 1950, 11, 147-154.

59. OSS ASSESSMENT STAFF. *Assessment of men.* New York: Rinehart, 1948.

60. OTIS, J. L. The effectiveness of the selection interview in appraising personality of salesmen. *Amer. Psychologist*, 1953, 8, 468. (Abstract)

61. PEPINSKY, H. B., SIEGEL, L., & VANATTA, A. The criterion in counseling: a group participation scale. *J. abnorm. soc. Psychol.*, 1952, 47, 415-419.

62. SAKODA, J. M. Factor analysis of OSS situational tests. *J. abnorm. soc. Psychol.*, 1952, 47, 843-852.

63. STOGDILL, R. M. Personal factors associated with leadership: a survey of the literature. *J. Psychol.*, 1948, 25, 31-75.

64. STOLPER, R. An investigation of several problems concerning counseling. Unpublished master's thesis, Louisiana State University, 1953.

65. TAFT, R. Use of the "group situation observation" method in the selection of trainee executives. *J. appl. Psychol.*, 1948, 32, 587-594.

66. TAFT, R. Some correlates of the ability to make accurate social judgments. Unpublished doctor's dissertation, Univer. of California, Berkeley, 1950.

67. VAN ZELST, R. H. Empathy test scores of union leaders. *J. appl. Psychol.*, 1952, 36, 293-295.

68. VERNON, P. E. The validation of civil service observation method in the selection of trainee executives. *Occup. Psychol.*, 1948, 32, 587-594.

69. VERNON, P. E. The validation of civil service selection board procedures. *Occup. Psychol.*, 1950, 24, 75-95.

70. VERNON, P. E., & PARRY, J. B. *Personnel selection in the British Forces.* London: Univer. of London Press, 1949.

71. WEISLOGEL, R. L. The development of situational performance tests for various types of military personnel. *Amer. Psychologist*, 1953, 8, 464. (Abstract)

72. WURSTER, C. R., & BASS, B. M. Situational tests: IV. Validity of leaderless group discussions among strangers. *Educ. psychol. Measmt*, 1953, 13, 122-132.

THE ATTENUATION PARADOX IN TEST THEORY[1]

JANE LOEVINGER

Washington University[2]

THE PARADOX

In recent years several authors have called attention to a paradox in test theory. Gulliksen (7) appears to have been the first. He showed that, under reasonable conditions, "In order to maximize the reliability and variance of the test, items should have high intercorrelations, all items should be of the same difficulty level, and this level should be as near to 50% as possible." But, he continued, "The criterion of maximizing test variance cannot be pushed to extremes. Test variance is a maximum if half of the population makes zero scores, and the other half makes perfect scores. Such a score distribution is not desirable for obvious reasons, yet current test theory provides no rationale for rejecting such a score distribution" (7, pp. 90–91).

All studies to be reviewed in this paper assume homogeneous tests, in the sense that all correlations between items within a test are accounted for by a single common factor. Validity, as used herein, refers to correlation of the test with that common factor. While the studies are explicitly concerned with reliability, in this context reliability refers only

to degree of homogeneity. There is no concern with stability of functions over time.

Tucker stated the paradox as follows: "Consider the case when all the items in a test are equivalent; that is, when the items all measure the same trait, have equal reliabilities, and are equally difficult. In this case the items are equally intercorrelated with coefficients equal to the item reliabilities . . . . If the reliability of the items were increased to unity, all correlations between the items would also become unity and a person passing one item would pass all items and another failing one item would fail all items. Thus the only possible scores are a perfect one or one of zero" (19, pp. 1–2). He pointed out, as a consequence of this paradox, that under these circumstances increasing the reliability of a test beyond a certain point will decrease the validity of the test, in contradiction to the usual belief, embodied in the correction for attenuation, that increasing the reliability of the test always increases its validity.

Brogden conceived of the problem as one of "determining the distribution of item difficulties which will maximize the correlation of the test with a perfect measure of the characteristic the test is intended to measure" (1, p. 197). With perfectly valid items, he pointed out, the difficulties should be equally spaced in some sense, whereas with items that do not intercorrelate, the difficulties should all be as close to .5 as possible. Brogden stated that item selection procedures aimed solely at increasing the

reliability of the test may result in a decrease in validity, again, the "attenuation paradox."

Two writers who missed the attenuation paradox were Loevinger (11) and Cronbach (3). Loevinger called attention to the anomalous result of using equivalent items when the item intercorrelations approached unity and concluded that equivalent items were undesirable in the usual case. Since the phi coefficient and KR 20 (Kuder and Richardson's [10] formula 20) measure the departure of the items from equivalence as well as the interrelationships of the underlying variables, she concluded that they were inappropriate as item-selection coefficients.

Cronbach, in refuting Loevinger, stated: "The phi coefficient which tells when items do and do not duplicate each other is a better index *just because* it does not reach unity for items of unequal difficulty" (3, p. 329). Here Cronbach neglected the fact that if one uses the phi coefficient for item selection, one needs two rules: For lower values of phi, the higher the coefficient, the more will the two items contribute to the validity of the test. But for high values of phi, the lower the coefficient, the more will the two items contribute to the validity of the test.

Similarly, Cronbach showed that the maximum value of KR 20 is not much less than unity for items with a specified distribution of item difficulties, and that the maximum value will drop for a greater range of item difficulties. If, however, maximizing KR 20 is made a rule for item selection, as Cronbach recommends, there will be a tendency to select items with a narrower range of difficulty, or, in Cronbach's terms, more redundant items. Here Cronbach apparently failed to see the paradox that whereas maximizing KR 20 for constant number of items will lead to increasing the validity of tests where the item intercorrelations are low, it will lead to decreasing the validity of tests where the item intercorrelations are high.

The resolution of the attenuation paradox thus lies in having two rules for test construction. For the "classical region," the region in which the attenuation of validity decreases with increase in reliability, the closer the items are to difficulty of .5 and thus to equivalence, the more reliable and more valid will the test be. For the "region of paradox" the optimal distribution of item difficulties must be determined as a function of item intercorrelations. This solution was implied by Brogden and stated by Davis (5).

DEFINITION OF THE REGION OF PARADOX

Four studies have contributed to the definition of the region of paradox, those of Tucker (19), Brogden (1), Davis (5), and Cronbach and Warrington (4).

Tucker (19) assumed that ability or true score is normally distributed, that probability of success on any item is related to true score by the normal ogive, and that all items are equally difficult. He found that with median equivalent items, i.e., equivalent items of difficulty equal to .5, the validity of the test constantly increased as the item reliability increased for a one-item test. For a 10-item test optimal item reliability (interitem correlation) was about .5, and for a 100-item test optimal item reliability was about .25. The maximum validity was .8 for a one-item test or for a test with perfectly reliable items, however many, since the

latter case is equivalent to that of one item. Maximum validity was .97 with 100 items and slightly over .9 with 10 items.

Tucker investigated a single case of non-median equivalent items. All items were of such difficulty that the ability level where the probability of passing is .5 was one standard deviation above the mean. (It will be convenient to refer to this case as tests for which $s = 1$. Results are identical for $s = -1$, that is, for tests where all items are of such difficulty that the ability level where half the individuals pass the item is one standard deviation below the mean.) In this case the optimal interitem correlation for a 10-item test was about .3, for a 100-item test about .15. The corresponding validities were about .83 and .96. For a one-item test the validity again increased as a function of item reliability to a maximum value of about .66. This value is also the terminal validity for a test of any number of items as the item reliability approaches unity, given that all items are of the specified difficulty level.

The optimal values of item reliabilities under his conditions were surprisingly low, Tucker pointed out; however, he cautioned that exceeding these values caused less decrement in validity than falling short of them. Further, where item difficulties are not all equal, his results do not hold exactly.

Brogden (1) assumed that the true score of ability was normally distributed, that the tetrachoric correlations of all pairs of items within a test were equal, and that the biserial r of each item with true score was the same for all items in a test. Phi coefficients between items and point biserials between items and true score were not always constant within a test because items were permitted to differ in difficulty. Validity coefficient of the test was studied as a function of tetrachoric r of items, number of items, and distribution of item difficulties. Item inter-r's had the values .2, .4, .6, and .8. Numbers of items were 9, 18, 45, 90, and 153. All item difficulties were concentrated at half sigma units between and including 2.0 and −2.0. The types of distributions were rectilinear, normal, skewed, and constant at each of the sigma values.

Some of Brogden's results are displayed in Fig. 1, 2, and 3. Figure 1 shows validity as a function of item intercorrelation for tests composed of median equivalent items, that is, items for which $s = 0$. The four curves correspond to four values for n, the number of items. Figure 2 is similar to Fig. 1, except that all items are of difficulty $s = 1$. (The curves of Fig. 2 also apply if all items are of difficulty $s = -1$.) While Fig. 2 again shows the attenuation paradox as a function of number of items, comparison of Fig. 1 and 2 shows the attenuation paradox as a function of item difficulty. Figure 3 shows validity as a function of distribution of item difficulties for tests composed of 90 items. The curve for $s = 0$ is also a member of another family of curves shown in Fig. 3. These curves show
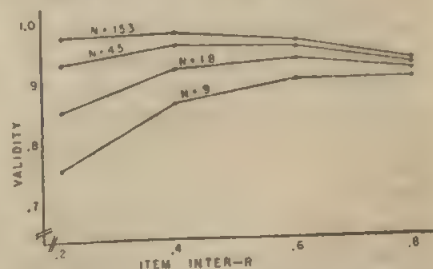
FIG. 1. A TTENUATION PARADOX AS A FUNC-TION OF NUMBER OF ITEMS FOR TESTS COM-POSED OF MEDIAN EQUIVALENT ITEMS. DATA FROM BROGDEN'S (1) TABLE 2.

what happens to a test of 90 equivalent items as the difficulty departs from median value. The data for these graphs are taken from Brogden's Tables 2 and 3.

Reading across the three figures, one sees that, in general, validity increases with increasing item inter-r up to a point and then decreases. The only curve which continues to rise in these figures is that for rectangular distribution of item difficulties. For a large number of nonmedian equivalent items, the optimal item inter-r is apparently less than .2. Paradox is exhibited whenever validity decreases as reliability (item inter-r) increases, that is, whenever the curve slopes downward. Brogden's method, unlike Tucker's, does not permit ascertaining exactly the optimal value of item inter-r, so the computed points have been connected by straight lines rather than attempting to sketch a curve. Tucker's article contains actual curves which may be compared with the figures here.

Reading up and down Fig. 1 and 2, one sees that validity always increases with increasing number of items, provided item inter-r and difficulty distribution are held constant; however, the increase in validity obtained by increasing n becomes smaller as the item inter-r increases.

Figures 1 and 2 show that optimal item inter-r decreases as n increases for constant item difficulty; in other words, the region of paradox increases as n increases. Comparison of Fig. 1 and 2 illustrates that validity drops as item difficulty departs from the median value for constant n and item inter-r. The same comparison illustrates that for tests composed of equivalent items, the region of paradox increases as the item difficulty departs from the median (Support for these generalizations is contained in further data not reproduced here.)

The two highest curves of Fig. 3 show an important manifestation of the attenuation paradox: Median equivalent items produce tests of higher validity for low values of item inter-r, while a rectangular distribution of difficulty produces tests of higher validity for high values of item inter-r. The curve for normal distribution of item difficulties lies between that for rectangular distribution and that for $z=0$ but was omitted to make the figure more legible. The point at which rectangular distribution produces higher validity than median equivalent items may
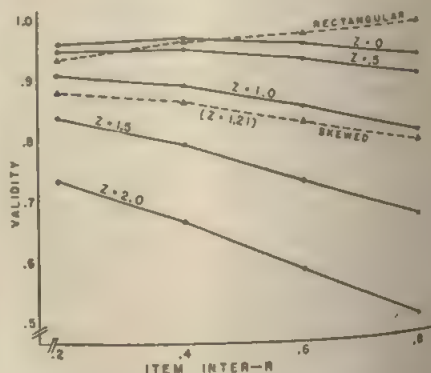
be taken as defining the limits of the region of paradox. Utilizing data from Brogden's Table 2 shows the following: Higher validity was obtained by tests of median equivalent items than by tests with distributed difficulties for item inter-r's of .2 and .4; however, for r of .4 and large number of items (90 or 153), there was almost no difference in validity for tests with median equivalent items and tests with item difficulties distributed normally or rectilinearly. For item inter-r equal to .6, the test with median equivalent items was superior for a 9-item test, ran about equal with tests of rectilinear and normal distribution of item difficulties for 18- and 45-item tests, and ran slightly behind those with distributed difficulties for 90 and 153 items. For item inter-r's equal to .8, distributed difficulties were clearly superior to equivalent items, more so as the number of items increased. The rectilinear had a slight advantage over the normal distribution for tests with more than 18 items.

The curve in Fig. 3 for skewed distribution of item difficulties falls far below the curves for rectangular, normal, and median equivalent items. This decrement is chiefly a result not of skewness but of the departure of mean item difficulty from $z=0$. The rectangular and normal distributions do have means at $z=0$, whereas for the skewed distribution the mean (computed by the reviewer) is at $z=1.21$.

For a given item inter-r, where tests are composed of items of constant difficulty, the validity falls off fairly sharply with the departure of that difficulty from the median; moreover, the greater the departure of the difficulty from the median, the lower the optimal value of the item inter-r. Comparing now the curve for skewed difficulties, which corresponds to a

mean $s=1.21$, with tests of constant difficulty equal to 1 and to 1.5, one obtains at least a hint that distribution of item difficulties protects the test from decrement of validity due to departure of difficulties from the median, at least for moderate and high values of item inter-r. More striking is the fact that despite the distribution of item difficulties, validity constantly falls with increasing item inter-r within the range considered, when the mean difficulty of the items is not appropriate for the group tested and the number of items is large. Practical considerations often lead to the use of a test with a group whose mean ability is not exactly the same as the mean difficulty of the items. Thus this important effect deserves much more extended investigation.

Davis (5) assumed that the most desirable distribution of test scores is a rectangular one and sought the conditions under which such a distribution could be obtained. He concluded that where the tetrachoric item inter-r's are .5 or less, the closest approach to a rectangular distribution will be obtained with median equivalent items. As the item inter-r rises above .5, the item difficulties should be increasingly dispersed. In comparing Davis' result with those of Tucker and Brogden, note that the latter investigators assumed that the true scores were normally distributed. Validity coefficients thus could equal unity only if obtained scores also were normally distributed. In effect, then, Tucker and Brogden assumed a normal distribution as a desideratum while Davis assumed a rectangular distribution.[3]

Cronbach and Warrington (4) assumed that the probability of success

[3] Davis' derivation, obtained from him in mimeographed form, appears to be somewhat less rigorous than the other two.

on any item is described by the normal ogive; however, in contrast to the preceding studies, in which persons with no ability were assumed to have zero probability of passing the item, they dealt only with three-choice items. Thus persons with no ability had a ⅓ chance of passing any item. The standard deviation of the ogive, σ_d, was the same for all items within a test. Most of the tests had 30 items. The underlying ability was assumed to be normally distributed. The scale value of each item is that level of ability for which the probability of passing the item is ⅔ (or, if corrected for chance, ½). Items were located at scale values between 2.5 and −2.0, inclusive, at intervals of .5. Pattern A had 30 items at scale value 0; pattern B had 6 at .5, 18 at 0, and 6 at −.5; pattern C had 6 each at 1.0 to −1.0 inclusive; and pattern D had 3 each at 2.5 to −2.0 inclusive. Cronbach and Warrington's chief problem was outside the realm of the attenuation paradox. They were concerned with the screening efficiency of the tests for various possible cutting scores from close to zero to over 90%. Screening efficiency was measured by r_{bis} of the dichotomized score scale against the continuous, normally distributed scale of ability. For each test a series of such validity coefficients was obtained and its curve plotted.

For perfectly precise items ($\sigma_d = 0$) the validity of pattern A was greatest only in the neighborhood of a cutting score of .5, and fell sharply below that of other patterns as the cutting score deviated from the median point. The range of cutting scores for which pattern A was superior increased as precision decreased until, for values of $\sigma_d \geqq 1$, pattern A was superior throughout most of the

range. A somewhat similar family of curves was generated by holding pattern constant and letting precision vary; that is, the greater the item precision, the more valid the test for cutting scores near the median, but the less valid for extreme cutting scores. This tendency was particularly clear for pattern A, which they called a "peaked" test, and is a manifestation of the attenuation paradox.

Cronbach and Warrington proposed an interesting integration of their findings. They suggested that increasing the variance of the scale values of the items (σ_v^2) has about the same effect as increasing the variance of the item ogive (σ_d^2), and that the important quantity to be considered is the sum of the two variances. Using eta rather than r as a measure of over-all validity, they found that under their conditions validity increased as a function of $\sigma_d^2 + \sigma_v^2$ to a maximum where the sum of the variances was about .5, then slowly declined.

At one point they stated, "Insofar as we can judge from these and Tucker's results, the peaked test has superior validity for even lower values of σ_d (higher r_{ij}) when the test is longer than thirty items" (4, p. 146). Cronbach and Warrington's data on the effect of test length are scanty, but the data of Tucker and Brogden are clear, and opposite to what Cronbach and Warrington indicate: The longer the test, the greater the region of paradox. (See Fig. 1 and 2.)

An incidental finding of Cronbach and Warrington's study was that the peaked tests tended to have bimodal or markedly skewed distributions, while the tests with dispersed difficulties had more nearly normal distributions. They did not consider the possibility that this finding is

... a manifestation of the attenua-
... paradox. As reason for using
... been stated, "I make particu-
moment r. r_{tt} is independent of the
test score metric. Our results are
therefore invariant as test scores are
transformed to other scales" (4, p.
142. But one may ask, what opera-
tions correspond to transforming test
scores to other scales? Given their
conditions, only their scale results.
The effect of various changes in
conditions would have to be investi-
gated. Indeed, it is generally con-
sidered a requirement for use of r_{bis}
that the dichotomized variable be
normally distributed. The effect of
using r_{bis} and eta as opposed to r
apparently is to increase the ad-
vantage of peaked as opposed to
distributed item difficulties, and thus
in effect to underestimate the region
of paradox.

In summary, the region of paradox
is most clearly defined in the studies
of Brogden and Tucker. Neither
Davis' study nor that of Cronbach
and Warrington detected that the
region of paradox increases with the
number of test items, an effect clearly
discernible in the two previous stud-
ies. If Cronbach and Warrington
had used the product-moment r in-
stead of eta they apparently would
have obtained results comparable to
those of Brogden and Tucker. Cron-
bach and Warrington called attention
to the importance of considering
items where there is a nonnegligible
probability of success without abil-
ity. Their method appears better
adapted to further exploration of this
problem than other methods, pref-
erably with r used as evaluation
function. A valuable extension of
their study would include two-choice
and four-choice items as well as ones
with no probability of success without
ability. Brogden's method has pro-

vided the most detailed information
concerning the region of the region of
paradox and would suggest to another
investigation of the relative merits
of concentrated and distributed diffi-
culties under circumstances where a
test may be used for groups differing
in mean ability. Cronbach and
Warrington's method could also be
used for the latter problem. Tucker's
paper, being analytic, is likely to
prove the most substantial contribu-
tion to test theory of any reviewed
herein.

APPLICATION TO TEST
CONSTRUCTION

The statistical phenomena sum-
marized by the phrase "the attenua-
tion paradox" reveal the importance
of the mean and variability of the
standardization sample in the evalua-
tion of a test's validity. Reading up
and down Fig. 3, one can interpret
the curves for differing values of s
as showing a striking decrement in
validity for tests composed of median
equivalent items when they are
subsequently applied to groups hav-
ing a different mean on the trait
measured. Similarly, the validity co-
efficients obtained for a hypothetical
set of tests with decreasing disper-
sion of item difficulties correspond
to those obtained by considering the
items as constant but administered
to groups of increasing variability.
Generally speaking, the standardiza-
tion sample should have the same
mean and variance as the group on
which the test will ultimately be
used. In practical situations usually
no such identity can be guaranteed,
certainly not for the life expectancy
of a well-constructed test.

For tests in the low homogeneity
or "classical region," as the variance
of the sample increases, mean held
constant, the validity coefficient in-

creases. Thus, to obtain a lower bound to the validity coefficient, the standardization sample should be at most as variable as any sample on which the test will be used.

For tests in the "region of paradox" the situation is more complicated. For a given degree of item intercorrelation and a given number of items, there is probably an optimal distribution of item difficulties. The higher the item inter-r, the more it is true that each item added to the test adds validity only if it differentiates at a different level of difficulty than those items already included. Any considerable increase in the variability of the group will decrease the validity of the test. Thus, to obtain a lower bound for the validity coefficient, the standardization sample should be at least as variable as any sample on which the test will be used.

Variation in the mean of the application sample from that of the standardization group must also be considered. Cronbach and Warrington concluded that where general usefulness over a period of time is sought items should be concentrated around median difficulty; however, they considered tests as selective instruments, which is legitimate but different from the present consideration of tests as measuring instruments, and their use of biserial correlation is questionable. Whether concentration of item difficulties makes a test more susceptible to loss of validity in application to groups of different means is a point on which evidence is not yet available.

Theoretical and practical considerations may be drawn together as follows: When application to a single group is considered, concentration of item difficulties is called for in the case of item inter-r's usually met, which are low. Consideration of use

of tests for differing groups will lead to dispersion of item difficulties for some cases, i.e., where item inter-r's are not too low. In practice, however, it is usually not possible to find large numbers of items exactly at the 50-50 point. The consequent inevitable dispersion of difficulties is in many cases probably desirable. A method of item selection that favors median items but does not exclude good items with more extreme difficulties seems indicated by these considerations. One such method has been proposed recently (13), and no doubt other methods have this property. That one widely used method of item selection has a very different effect will be shown in the next section.

IMPLICATIONS FOR TEST THEORY

Consider the definition: A *paradoxical property* of a test is a property such that the validity of the test is not a monotonic function of that property; i.e., validity is sometimes an increasing and sometimes a decreasing function of the property.

The objections to taking validity as the focal concept for test theory are well known. Probably all psychometricians would agree that test theory must have basic concepts which refer to intrinsic properties of tests. Is it not intuitively valid, however, to demand that the most basic concept of psychometrics shall be a nonparadoxical property of tests? Are there such nonparadoxical properties?

Clearly reliability, or one of its cognate concepts, is the focus of present-day test theory; the statistical theory of reliability is the bulk of classical test theory. There have been many criticisms of the reliability concept; those by Thorndike (17), Cronbach (2), and Loevinger (11) have a good deal in common, particu-

larly the distinction between stability and homogeneity, which is lost in the ordinary usage of "reliability." The purpose of the present article is not, however, to review all the objections to the reliability concept but to draw attention to a single instance in which the statistical theory of reliability leads to self-contradiction.

Such solutions and explanations of the attenuation paradox as have been proposed, those of Brogden, Davis, and Cronbach and Warrington, are appeals to common sense completely outside traditional theory of reliability. Perhaps the most paradoxical aspect of the attenuation paradox is that Gulliksen, who appears to deserve credit for discovering it, failed to include any reference to it in his comprehensive summary (8) of mental test theory. In his own words, "Current test theory provides no rationale for rejecting" the anomalous score distribution.

Lord (14) has probably come closest to integrating the attenuation paradox with classical test theory. He points out that while Pearsonian correlation between test score and ability decreases in what is here called the region of paradox, the curvilinear correlation with ability constantly increases with increasing item inter-r. Unfortunately, however, consideration of curvilinear correlation obscures the paradox, which does in fact exist. Lord concerns himself chiefly with the classical region; it is not clear that his approach could lead to the distinction between the two regions of test theory.

Reliability is paradoxical; saturation, a concept imbedded in a method of test construction recently proposed by Loevinger, Gleser, and DuBois (13), is virtually identical with one of the Kuder-Richardson (10) co-

efficients and thus is paradoxical in the same region. Guttman's (9) concept of scalability and Loevinger's (11) concept of homogeneity clearly differ from reliability and saturation. Are these properties paradoxical? Consideration of the results quoted in the second section of the present paper shows that these properties are also paradoxical, not, however, in what is here called the region of paradox but in the classical region. To distinguish the two kinds of paradoxical properties, scalability and homogeneity may be called neoparadoxical properties. Only scale theory will be considered. Its importance is pointed up by the fact that virtually all recent contributions to test theory in sociology and closely related disciplines are phrased in terms of scale analysis.

Guttman originally measured scalability in terms of a "coefficient of reproducibility," which is the proportion of the responses of the group which can be reproduced from knowledge of item difficulties and total scores. Guttman distinguished "scales" from "quasi-scales" according to whether the coefficient of reproducibility exceeded .9, assuming certain other conditions were satisfied. Festinger (6), Loevinger (12), and probably others criticized the distinction between scales and quasiscales as arbitrary. According to these writers, Guttman and his followers were erecting a qualitative difference out of a purely quantitative one. The considerations of the present paper, however, justify what was apparently a purely intuitive distinction on the part of Guttman. When we deal with cumulative (12) dichotomous items, which is the usual case (16), the distinction between the two regions of test theory can be clearly, if laboriously, drawn. What

Guttman called quasi-scales are tests in what is here called the classical region; what Guttman called scales are tests in what is here called the region of paradox. There is no evidence, however, that workers in the field of scale analysis are aware of the precision with which this distinction can be made, nor of the consequences of the distinction which have been elaborated by psychologists working in test theory.

A number of objections have been raised against the original coefficient of reproducibility, a major one being that its lower limit can be arbitrarily raised by selecting items extreme with respect to difficulty (or popularity). In practice, workers in this field have probably always more or less taken this fact into consideration. A recent paper by Menzel (15) has proposed a modification of the coefficient of reproducibility which eliminates this problem. Clearly, insofar as use of Guttman's coefficient led to the selection of items extreme in difficulty as opposed to median items, it led to a decrease in validity of the resultant test, since under none of the conditions investigated above did a selection of extreme rather than median items lead to an increase in validity.

Stouffer, Borgatta, Hays, and Henry (16) report that, in practice, in order to derive scales rather than quasi-scales from available data, it has been necessary to select one from among several apparently equally good items at any given difficulty level. According to them, probably the most common method of improving scalability has been reduction of number of items per test, often to no more than four or five items. Inspection of Fig. 1 and 2 and other available data reveals no set of conditions under which reduction in number of items increases validity;

on the contrary, increase in number of items invariably increases variance most markedly when the number of items is small.

A further consequence of using the concept of perfect scales as the desideratum in test construction has been that test constructors have been forced to select items that are more or less evenly spread in difficulty. Such selection will tend to increase scalability, but in the classical region it will tend to decrease validity.

In summary, Guttman's coefficient of reproducibility gives advantage to extreme as opposed to median items. If this coefficient is modified so as to give advantage neither to extreme nor to median items, pursuit of scalability still leads to decrease in number of items and to dispersion of item difficulties. But the evidence cited in this review is that decrease in number of items always leads to decrease in validity, other things equal; there is no set of conditions under which extreme items lead to more valid tests than median items, and in the classical region, dispersion of item difficulties leads to decrease in validity. There is reason to believe that most of the attitude tests currently in use fall in the classical region. One may conclude that extensive use of scale analysis has almost certainly led to loss of validity in tests used in sociology.

In recognition of the difficulties to which scale analysis has led, more or less akin to those cited here, Stouffer, *et al.* (16) have proposed a modification of the method. They continue to have about five items per scale, but each item is a "contrived item." A contrived item is a composite of several items similar in level of difficulty, but each contrived item is scored zero or one, depending on the number of pluses in the items of

which it is composed. The contrived items are constructed by a cut and try method. The authors show that tests constructed by their method are superior to those constructed by the more common method of scale analysis with respect to scalability and some definition of reliability, but only where the original items are quite good. Their method appears to be a compromise between scale analysis and methods favoring choice of equivalent items, to which most psychologists would lean. While the new method may ameliorate the flaws of scale analysis, it does not solve the conceptual difficulty. Scalability, like reliability, is a paradoxical concept. Worse, scalability is paradoxical in the classical region, which is heavily populated with tests, while coefficients such as the Kuder-Richardson (10) reliability coefficients are paradoxical only in the region of paradox, which is thinly populated with tests.[4]

The problem of finding nonparadoxical properties of tests remains. A property different from those discussed above which may qualify is the discriminating power of the test. Many writers have used this term without definition, as if it were self-explanatory. Several recent papers have provided indices of discriminat-

ing power. ... demonstration of these ... and be stood. A sequence ... is selected arbitrarily to show that this concept differs from concepts like reliability and scalability.

In a paper Gleser and DuBois [13] have distinguished three aspects of discriminating power: fineness, probability, and range. Discriminating fineness refers to the size of the differences in the trait which can be discriminated. Discriminating probability refers to the proportion of discriminations which are in the same direction as trait differences. Discriminating range refers to the general level of the trait at which discriminations are made. No single coefficient measures the three aspects of discriminating power; however, some coefficients are closely related to discriminating fineness, some to discriminating probability, and some to both. If the test construction process is conceived as beginning with the best few items in a finite pool and adding items one at a time from the pool in the order of the goodness of the items, then coefficients measuring only fineness constantly increase, those measuring only probability constantly decrease, while those measuring both may increase at first and then decrease. The latter type of coefficient provides a basis for deciding when to stop adding items to the test.

This definition of discriminating power may prove unappealing to many psychometricians, since no single coefficient corresponds to it and since it is essentially an intuitive rather than a quantitative concept. Yet other quantitative disciplines begin frankly with intuitive concepts. As Ruth Tolman (18) has recently observed, among physical scientists it is the highest compliment to speak

---

[4] Dr. Ledyard Tucker has, however, called my attention to the fact that a problem being worked on by John Keats at Princeton University provides a mathematical model for the method of "contrived items." Keats assumes that the probability of success on an item increases with ability at only one point on the scale of ability, rather than describing an ogive. It is not assumed that the probability of success on the item is zero below that point nor unity above it, only that it is constant everywhere but at the single point. One has difficulty thinking of test content for which this assumption is as reasonable as the assumption of a graded increase in the probability of success on items.

of a fellow scientist as having "physical intuition," but among those psychologists who most desire to emulate the physical scientists, intuition is rarely referred to in a similar complimentary sense.

In summary, paradoxical property of a test is a property such that validity is not a monotonic function of that property. Reliability is a property paradoxical in a region lightly populated with tests. Scalability is a property paradoxical in the region heavily populated with tests. Other possible properties, such as the discriminating power of the test, have not been fully investigated.

REFERENCES

1. BROGDEN, H. E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. *Psychometrika*, 1946, 11, 197–214.
2. CRONBACH, L. J. Test "reliability": its meaning and determination. *Psychometrika*, 1947, 12, 1–16.
3. CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297–334.
4. CRONBACH, L. J., & WARRINGTON, W. G. Efficiency of multiple-choice tests as a function of spread of item difficulties. *Psychometrika*, 1952, 17, 127–147.
5. DAVIS, F. B. Item analysis in relation to educational and psychological testing. *Psychol. Bull.*, 1952, 49, 97–121.
6. FESTINGER, L. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, 44, 149–161.
7. GULLIKSEN, H. O. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, 10, 79–91.
8. GULLIKSEN, H. O. *Theory of mental tests.* New York: Wiley, 1950.
9. GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 9, 139–150.
10. KUDER, G. F., & RICHARDSON, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, 2, 151–160.

11. LOEVINGER, JANE. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, 61, No. 4 (Whole No. 285).
12. LOEVINGER, JANE. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychol. Bull.*, 1948, 45, 507–529.
13. LOEVINGER, JANE, GLESER, GOLDINE C., & DUBOIS, P. H. Maximizing the discriminating power of a multiple-score test. *Psychometrika*, 1953, 18, 309–317.
14. LORD, F. M. A theory of test scores. *Psychometr. Monogr.*, 1952, No. 7.
15. MENZEL, H. A new coefficient for scalogram analysis. *Publ. Opin. Quart.*, 1953, 17, 268–280.
16. STOUFFER, S. A., BORGATTA, E. F., HAYS, D. G., & HENRY, A. F. A technique for improving cumulative scales. *Publ. Opin. Quart.*, 1952, 16, 273–291.
17. THORNDIKE, R. L. Reliability. In E. F. Lindquist (Ed.), *Educational measurement.* Washington: American Council on Education, 1951. Pp. 560–620.
18. TOLMAN, RUTH S. Virtue rewarded and vice punished. *Amer. Psychologist*, 1953, 8, 721–733.
19. TUCKER, L. R. Maximum validity of a test with equivalent items. *Psychometrika*, 1946, 11, 1–13.

REGRESSION ANALYSIS: PREDICTION FROM CLASSIFIED VARIABLES[1]

ROBERT M. GUION

Bowling Green State University

Occasionally in psychological research a qualitative independent variable seems promising in the prediction of a quantitative variable. Sometimes, particularly in applied research, data for basically continuous independent variables must be obtained, if at all, in terms of broad, discrete categories, possibly of unequal intervals. Such variables are at best inconvenient to work with in prediction problems; all too often, the difficulties they present are solved by ignoring them, resulting in the loss of a potentially useful predictor.

In a study of supervisors (2), the writer was faced with this problem. It was necessary to develop a regression equation by which a dependent variable (number of employees supervised) could be predicted from such independent variables as man-hours of assistance, rough classifications of plant size, and the purely qualitative variable of industry classification. The problem was solved by a technique of regression analysis.

Regression analysis is a least-squares regression technique permitting the prediction of a quantitative variable from categorized or qualitative variables.[2] There seems

to be no particular limit to the number of variables that can be handled by this method, as there is in the usual multiple regression techniques. Apparently the absence of products in the resulting regression equation reduces the likelihood of cumulating to a serious degree the errors of measurement. It is, of course, possible that spurious correlation might result from including too many variables.

It is beyond the intent of this report to give a complete account of the theory and practice of regression analysis. The procedure, which has been used in agricultural research, is capable of further refinement and needs evaluative research. A description of the basic theory and of the method's application in agriculture is provided by Anderson and Bancroft (1). This report seeks merely to outline the procedure as modified in the writer's use of it.

The technique will be discussed and outlined by example, using a three-variable situation that might occur in personnel research. For purposes of illustration, we will be concerned with predicting the average number of working days attendance per quarter year of machine operators. The prediction will be based upon a knowledge of the applicant's categorical standing in each of three variables:

Variable A: Preference for mechanical work

1. First or second choice
2. Third choice or not indicated

Variable B: Housing status

1. Own home
2. Rent home
3. Rooming

Variable C: Presently employed?

1. Yes
2. No

There are seven categories in all in this example; any applicant may be classified as being in three of them. Regression analysis seeks to estimate eight parameters: α_1, α_2 for the two categories of variable A; β_1, β_2, β_3 for the three categories of variable B; γ_1, γ_2 for the two categories of variable C; and μ for the mean attendance (dependent variable) of the general population of applicants. The sum of the parameters for the categories describing any individual, when added to the population mean, yields the dependent measure for that individual within limits of error, such that the regression model is

$$x = \alpha_i + \beta_j + \gamma_k + \mu + \text{error}.$$

From a sample available for study, we can estimate these parameters (estimates indicated by italic letters) so that

$$x \doteq a_i + b_j + c_k + m.$$

In other words, an adjustment factor is assigned to each category of each variable. For any individual applicant in the illustrative situation, the predicted attendance record would be equal, within limits of error, to the mean attendance record of the total sample plus the algebraic sum of the adjustment factors of the categories that describe him.

A inexpensive solution of these parameter estimates is sought by using a system of eight simultaneous equations (the number of categories plus one) by the following procedure.

1. *Prepare the matrix of the system of equations.* The equation for the first category of the first variable is

$$\sum X_{a_1} - n_{a_1} a_1 - n_{a_1 b_1} b_1 - n_{a_1 b_2} b_2$$
$$n_{a_1 b_3} b_3 - n_{a_1 c_1} c_1$$
$$n_{a_1 c_2} c_2 - n_{a_1} m + 0.$$

ΣX_{a_1} is the sum of the measures of the dependent variable (i.e., attendance in days) for all cases appearing in the first category of variable A: those who have indicated on the application blank a first or second choice for mechanical work. The frequency, or number of cases, appearing in this category is designated n_{a_1}. The notation $n_{a_1 b_1}$ indicates the number of cases in this category which are also classified in the first category of variable B: those who own their homes. The next six equations follow this form, being equations for the second category of variable A, the first category of variable B, and so on.

The final equation is for the general mean and is

$$\sum_{i=1}^{N} X_i - n_{a_1} a_1 - n_{a_2} a_2 - n_{b_1} b_1 - n_{b_2} b_2$$
$$- n_{b_3} b_3 - n_{c_1} c_1 - n_{c_2} c_2 - N m = 0.$$

Transposing all parameter estimates and their coefficients yields a matrix equal to the vector of the dependent variable. This 8 by 8 matrix of independent variables consists of a set of frequencies as coefficients of the unknown parameter estimates. The complete system of equations, in matrix form blocked off for illustrative purposes, is shown as Table 1.

It will be seen that the total fre-

quency in any given category appears in the diagonal, and that the coefficients of the diagonal are equal to the coefficients of the last equation for the general mean.

The sample used in this illustration co...ists of 49 machine operators, 26 of whom have indicated machine work either as their first choice or second (first category, variable A). Of these 26 men, 7 own their homes, 6 are renting, and 13 are rooming or living with their parents (the three categories of variable B, respectively). They totaled 1,417 working days attendance during a three-month period.

Blocking the matrix off as shown in Table 1 provides a convenient check on the accuracy of the tabulations made: the sum of the frequencies in any given block must be equal to the total number of cases in the sample unless there are individuals included for whom complete data are not available. In any column within a block, the sum of the coefficients is equal to the coefficient in that column in the equation for the general mean.

2. *Reduce the matrix.* This system of equations, as it now stands, is additive since the sum of the coefficients of the equations for each variable yields the appropriate coefficients for the equation for the general mean. This set of equations does not have full rank and therefore has no unique solution.

A solution can be obtained by making the restriction that the sum of the estimated parameters, or a weighted sum, be equal to zero for any given variable. With this restriction, the process of reparametrization, discussed by Kempthorne (3), can be applied. Reparametrization seeks to replace certain parameters and solve for a different set. It can

be done by setting one .... each variable equal to .... results in each case in the .... the last category of each variable (coefficients in Table 1, italicized), we have the following reduced 5 by 5 matrix of equations to solve instead of the original 8 by 8:

$$
\begin{vmatrix}
26 & 7 & 6 & 11 & 26 \\
7 & 13 & 0 & 10 & 13 \\
6 & 0 & 15 & 6 & 15 \\
11 & 10 & 6 & 23 & 23 \\
26 & 13 & 15 & 23 & 49
\end{vmatrix}
=
\begin{vmatrix}
1417 \\
743 \\
842 \\
1260 \\
2700
\end{vmatrix}.
$$

This reduced matrix can be solved.[4] The solutions for the illustrative problem are shown in Table 2.

3. *Solve the equations.* There are, of course, many methods of solving sets of simultaneous equations. However, in systems as large as will be found in most cases of regression analysis, loss of significant figures in the course of computation will become a serious problem if direct reduction or a determinant method is used. In the supervisory study cited, for a matrix of 27 equations, the Gauss-Seidel iterative method was used. The 134 iterations required to reach a solution were performed on the IBM Card-Programmed Calculator using a procedure described by Liggett (4).

TABLE 1

MATRIX OF COEFFICIENTS OF THE UNKNOWN PARAMETER ESTIMATES AND VECTOR OF DEPENDENT VARIABLE

Categories	Coefficients of Parameters								Vector (ΣX)
	a_1	a_2	b_1	b_2	b_3	c_1	c_2	m	
Variable A									
Category 1	26	*0*	7	6	*13*	11	*15*	26	1417
Category 2	*0*	23	*6*	*9*	*8*	12	*11*	23	*1283*
Variable B									
Category 1	7	*6*	13	0	0	10	*3*	13	743
Category 2	6	*9*	0	15	0	6	*9*	15	842
Category 3	*13*	*8*	0	0	21	7	*14*	*21*	*1115*
Variable C									
Category 1	11	*12*	10	6	*7*	23	*0*	23	1260
Category 2	15	*11*	3	9	*14*	*0*	26	26	*1440*
General Mean	26	23	13	15	21	23	26	49	2700

Note: Italicized values are those eliminated in reparametrization.

4. *Convert the obtained values.* The solution after reparametrization yields for the illustration the values a_1', b_1', b_2', c_1', and m', since the last category of each variable had been set equal to zero. Conversion of these estimates to estimates of the original set of parameters is accomplished by simple algebra, the actual process depending upon the restriction first imposed. One of three restrictions might have been made:

1. The sum of the estimated parameters (or adjustment factors) can be set equal to zero, if the population can be assumed to be equally distributed among the various categories of the variable.

2. The sum of the estimated parameters of each variable, weighted according to the observed frequencies of the categories, can be set equal to zero, if it is assumed that the sample is random.

3. The sum of the estimated parameters of each variable, given a priori weights on the basis of known or hypothesized population frequencies, can be set equal to zero.

The second restriction seems more common and will be used in the example. Following Kempthorne (3), it can be shown that $a_i' + a_k = a_i$. It is therefore necessary to find the kth value for each variable, i.e., the value eliminated in reparametrization. The restriction is that $\sum_i^k n_{a_i} a_i = 0$, or, in the example where variable B has three categories, $\sum_s^3 n_{b_i} b_i = 0$. Substituting, this becomes in full

$$n_{b_1}(b_1' + b_3) + n_{b_2}(b_2' + b_3) + n_{b_3}b_3 = 0.$$

TABLE 2

SOLUTIONS TO ILLUSTRATIVE EQUATIONS

Parameter	Prime Solution*	Parameter Estimate
a_1	−1.048	−0.492
a_2		+0.556
b_1	+4.918	+2.689
b_2	+3.020	+0.791
b_3		−2.229
c_1	−1.865	−0.990
c_2		+0.875
m	+54.316	+55.102

* Solution to reparametrized matrix.

Simplifying and solving for b_3,

$$b_3 = - \frac{n_{b_1}b_1' + n_{b_2}b_2'}{N},$$

or, in the illustration,

$$b_3 = - \frac{(13)(4.918) + (15)(3.020)}{49}$$

$$= - 2.229,$$

and,

$$b_1 = b_1' + b_3 = 4.918 - 2.229 = 2.689$$

$$b_2 = b_2' + b_3 = 3.020 - 2.229 = 0.791.$$

The same general procedure is used to obtain the correct parameter estimates, or adjustment factors, for variables A and C, shown in Table 2. For the general mean, it is best to use the simple arithmetic mean as normally computed because of the rounding errors which creep into the corrected value after the solution through reparametrization. In this problem, the mean is 55.102.

With these values known, the linear equation

$$x = a_i + b_j + c_k + m$$

can be used for each applicant. In the illustration, we can now, by knowing the applicant's classification in the three variables and by knowing the mean attendance of the sample, predict the applicant's job performance in terms of expected attendance per quarter.

For example, an applicant who lists mechanical work as his first job choice (category a_1, -0.492), who owns his home (category b_1, $+2.689$), and who is presently employed (category c_1, -0.990) would be expected to have a quarterly attendance record, in integral units, of 56 working days, the prediction formula being

$$x = 55.102 - 0.492 + 2.689 - 0.990$$

$$= 56.309.$$

The relative influence of each independent variable can be gauged according to the variance of the parameter estimates, such that, in this problem, variable B carries a heavier weight in prediction than variable C, which is in turn weighted more heavily than variable A.

It should be pointed out that this problem has been illustrative only; where the units are discrete rather than continuous, which is true of both the frequencies of individuals in various categories and of the measurements of attendance, there is probably very little justification in attempting to achieve greater than integral accuracy.

CONCLUDING COMMENTS

Regression analysis is a least-squares multiple regression technique permitting the prediction of quantitative variables from qualitative or classified variables. As outlined here, regression analysis assumes no significant interactions between variables. If interaction is found to exist, additional equations can be introduced, treating combinations of interacting variables as separate classifications. It should be pointed out, for the sake of economy, that any error made by a false assumption of no interaction is an error of underestimating rather than overestimating the relationship; the effect is one of ignoring additional variables which could have increased the predictive efficiency.

The regression equation derived from this method does not directly provide certain interpretational data. However, the predicted values of the dependent measure can be correlated by conventional procedures with the actual values. These correlation coefficients can then be used to derive coefficients of determination and standard errors of estimate.

The potential uses of the method

are many. Perhaps the most obvious are its applications to personnel history analysis and to profile analysis. Nevertheless, it should be made the subject of empirical investigations. Several research questions need answering. Its predictive effectiveness

in comparison with other procedures, such as the Horizontal Profile Method [5, p. 256], seems to be a requirement, as should the effect of varying groupings or of the number of variables used.

REFERENCES

1. ANDERSON, R. L., & BANCROFT, T. A. *Statistical theory in research.* New York: McGraw-Hill, 1952

2. GUION, R. M. The employee load of first-line supervisors. *Personnel Psychol.*, 1953, 6, 223-244.

3. KEMPTHORNE, O. *Design and analysis of experiments.* New York: Wiley, 1952.

4. LORGE, I. C. Two applications of the

IBM card programmed electronic calculator. *Proc. Inquiry Conference, Statistics, September 1950, 67-72.*

5. SEGAL, W. H., SCHMITT, C. J., & ASSOCIATES. *Occupations.* New York: American Book Co., 1948.

Received September 16, 1953.

ESTIMATING THE SCALABILITY OF A SERIES OF ITEMS—AN APPLICATION OF INFORMATION THEORY

RICHARD WILLIS

University of Minnesota

The powerful scaling technique developed by Guttman and others (3, 4, 5, 10) has several advantages, some of which seem to be unique. The merits and limitations of the method have been evaluated elsewhere (1, 2, 6), and they will not be reviewed here. The present paper will be restricted to a consideration of the problem of estimating the degree of scalability present in the area in question; i.e., the degree to which items in the scale are measuring the same thing in the population members.

Guttman discusses several factors which should be considered in this connection, namely, (a) the number of items, (b) the number of respondents, or size of the population sample, (c) the number of errors, or responses which do not fit the pattern of a perfect scale, (d) the distribution of these nonfitting responses, (e) the number of response categories of the items, and (f) the distribution of the item marginals. The first three of these criteria are used to compute a coefficient of reproducibility r, which is equal to $1 - p_E$, where p_E is the proportion of errors among the total number of responses. This coefficient should equal or exceed .90. The fourth factor is determined by an inspection of the response pattern. Nonfitting responses should be well scattered, indicating a random distribution, rather than appear in clusters, which would indicate a systematic distortion of the scale pattern. The last two criteria are ac-counted for by rule-of-thumb procedures. The more response categories given to the items individually, and hence the larger total number of response categories, the less likely is a scale pattern to appear by chance alone. It is recommended that the number of items be ten or more if they are all dichotomous, so that the number of response categories is at least twenty. Likewise it is recommended that there be included no more than a few items that have almost all responses lumped under a single alternative, as such items inevitably boost the coefficient of reproducibility. For example, an item to which 95 per cent of the respondents answer "agree" could not possibly result in more than 5 per cent nonfitting responses, while an item with a 50-50 split between two categories could theoretically result in 50 per cent errors. Items with more than two alternatives might produce even at best more than 50 per cent nonfitting responses if not removed from the series.

Because of the effects which the number of categories and the distribution of category frequencies have on reproducibility coefficients, they are not strictly comparable as generally computed. A means of quantitatively accounting for these effects would be especially useful because it is standard practice to administer each item with several categories, and then to combine adjacent categories in an attempt to get r up to .90. And, of course, the category

frequencies are changed at the same time. Another common practice in which these effects play an important part is that of selecting a subgroup of items from among those administered. In both cases the manipulation is selective, favoring the result sought by the experimenter, and should be corrected for.

The concept of *entropy*[1] has been applied to the theory of information by Wiener (11) and by Shannon (9). This interesting concept provides the means for such a correction. The entropy *H* associated with a contingent event is an indication of the number of possible outcomes it can have. It is consequently a measure of the uncertainty that the event will occur in any specified one of these possible ways. The toss of a die, with six equally likely possible outcomes, involves more entropy than the toss of a coin, with only two equally likely outcomes possible. Knowing the outcome of a high-entropy situation gives us more information, in the sense that more uncertainty is removed, than knowing the outcome of a low-entropy situation.

How does this apply to estimating the degree of scalability present in an area? If we make our estimation first under one set of conditions and then make a second estimation under a different set of conditions having a different associated entropy value, the two estimates are not comparable unless we can allow for this difference. This is true whether we have tested in the same area twice or in two completely different areas. The situation is somewhat analogous to comparing

two chi-square values without considering a difference in the degrees of freedom. Let us imagine, for example, that we have given ten treacherous items to 100 respondents. Then we arrange the responses so that they form as nearly as possible a scale pattern, using any of the available methods (3, 5, 10). Say the coefficient of reproducibility *r* turns out to be .80. This does not meet the accepted standard of .90, and we begin to judiciously combine adjacent categories so that we are left with ten dichotomous items and an *r* of .90. Is this really any stronger evidence that the area is satisfactorily scalable? The proportion of responses which do not fit the scale pattern has been cut in half, but perhaps the restructuring of the situation has reduced the likelihood of such nonfitting responses by as much or more. The more we manipulate the situation to our advantage, the less chance there will be for errors to show themselves. Our judicious manipulation reduces the entropy and imposes greater restrictions on the responses, forcing them to more nearly fit the scale pattern.

To see how much this chance for errors to appear has been reduced, let us calculate entropy values for conditions before and after combining categories. The formula is

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i, \qquad (1)$$

in which p_i is the relative probability of the event i occurring.[2] As p_i will always be less than 1, if there is any uncertainty whatever about the outcome, $\log_2 p_i$ will be negative and *H*

---

[1] It has been so named because of its mathematical and conceptual similarity to the term entropy used in statistical mechanics, which in turn was named after the entropy of classical thermodynamics. For another application of entropy-like considerations in psychology, see Miller and Frick (7).

[2] Shannon (9) discusses the characteristics which a good indicator of uncertainty should have and shows that this is the only form of expression which does have these characteristics.

will be positive. The logarithm is taken to a base of 2 only because H then takes on a convenient numerical value—it then indicates how many times the number of equally likely possibilities must be halved to remove all uncertainty about the outcome. An H value of 1.00 would represent an event such as the toss of a coin, in which the equally likely possibilities are reduced from 2 to 1 when the outcome becomes known. This same H value would also characterize a single respondent checking a dichotomous item, assuming we have no a priori knowledge about which category he is more likely to endorse, i.e., assuming $p_1 = p_2 = .5$. In either the case of the coin or the single respondent answering a single dichotomous item, formula 1 reduces to

$$H = -.5 \log_2 .5 - .5 \log_2 .5$$
$$= -.5(-1) - .5(-1) = 1.00.$$

But in our imaginary example above, we had 100 respondents instead of only one, and so we need not assume that all the presented choices for an item are equally likely to be chosen. We obtain an estimate of the relative probabilities from the response frequencies, and the larger the group, the better will be our estimates. The response frequencies give us a set of p_i's for each item such that $\Sigma p_i = 1.00$; using these p_i's, we can calculate an H value for each item in the series. Then we sum item H's, and the result is an H value characteristic of our series of items. This value indicates how much information is obtained (i.e., how much uncertainty is removed) each time we administer our item series to a respondent. The formula may be written

$$H = -\sum_{j=1}^{I} \sum_{i=1}^{c_j} p_{ij} \log_2 p_{ij}, \quad (2)$$

where p_{ij} is the relative probability of category i on item j being endorsed, I is the number of items in the series, and c_j is the number of response categories presented with item j. The summation is first over categories within items and then over items. Note that this formula does *not* take into consideration the patterning of the responses. It is the coefficient of reproducibility which does this, and that is why the two measures must be considered together in order to see the whole picture.

Returning once more to our example, suppose we compute the entropy before, H_1, and after, H_2, combining categories with the result that $H_1 : H_2 = 2$. Then the nonfitting responses have been reduced only in proportion to the reduction in entropy. We should conclude that the increase in r is most likely due to sheer manipulation. It should be pointed out here that increased reproducibility after combining categories can mean more than this. Such would be the case if respondents were differentiating between the first category on an item and the other two according to their true scale position, but were distinguishing between the second and third categories only by chance, or by habits of expression, or by any other extraneous factor. But before we could conclude that this sort of thing was happening, it seems reasonable that we should require that the errors, or nonfitting responses, be reduced as much as or more than the entropy, so that

$$p_{E_1}/p_{E_2} \geq H_1/H_2 \quad (3)$$

where p_{E_1} and p_{E_2} are the relative proportion of errors before and after combining categories, respectively. Lowering H means a greater structuring of the situation, and as we have been deliberately structuring it in

our favor, we must demand not only a reduction in errors, but a reduction at least in proportion to this increased structuring.

The proportion in formula 3 may also be written as

$$H_2/p_{E_2} \geq H_1/p_{E_1},$$

and this suggests the possibility of setting up a standard H/p_E ratio with which to compare H to p_E values obtained in practice. Taking into consideration the values of *both* r and H p_E would decrease the possibility of accepting as scalable an area which is not. It might also occasionally provide the evidence needed to accept as scalable an area about which we would have otherwise been in doubt. Choosing a standard ratio, call it H_s/p_{E_s}, is largely a matter of deciding how conservative we wish to be. It seems logical to choose $p_{E_s} = .1$, as this corresponds to an r of .90. We have a rough guide in choosing H, in Guttman's statement (10, p. 79) that at least ten items should be used if they are all dichotomies. If we assume that the p_i's are normally distributed, so that each p_i is associated with an equal portion of the area under the probability curve, H_s/p_{E_s} then becomes $8.9/0.1 = 89$. If we accept this standard ratio, then we would require an r of more than .90 should the entropy of the item series fall below 8.9. If, for example, our item series had an H value of 17.8 (which would be extremely high), an r of .80 would most likely be as strong an indication of scalability as the r of .90 in the first example, although a conservative experimenter would probably prefer to exceed the minimum values for both r and H/p_E.

A previously mentioned situation which frequently occurs in practice, for which the above considerations are also in order, is that in w··· subset of items which seems ··· a scale is chosen out of the ··· number of items administere···. Choosing such a subset impos··· greater structuring on the sit··· and produces a corresponding H drop. As in the case of combining cat··· ies, H/p_E as well as r should be no··· before drawing conclusions. Whet··· or not an experimenter chooses ··· use a standard H to p_E ratio, tha··· gested above or any other, it w··· seem advisable in any case to re··· any manipulation, such as comb··· categories or picking out sub··· items, and the resulting H drop··· along with the coefficients of re··· ducibility, as this information will ··· helpful in interpreting these r's.

When categories are combined or subsets of items are picked out ac- cording to the usual practice, there are actually two distinct processes going on simultaneously. First, there is the increased structuring and the concomitant decrease in entropy. Second, there is the process of selec- tion carried on by the experimenter in which he capitalizes on the oppor- tunities which are presented by the data. The first process is an auto- matic result of decreasing the number of items or response categories and would obtain no matter whether the combining of categories or selection of item subsets is done selectively or entirely at random. It is the first process, the automatic H drop, which is adjusted for by the H/p_E ratio, but there remains the problem of account- ing for the process of selection.

It would seem that the only test of scalability which allows for this second factor is that of cross valida- tion on a new sample. And obviously, the greater the H drop, the greater the possible influence of selection, and therefore the more esssential it be-

comes to cross validate. Not only ~~should the two coefficients of reproducibility, and the new H values taken~~ together, meet the minimum values, but it is also important that the items arrange themselves in the same order of difficulty or very nearly so. It is also desirable that the most advantageous way of combining response categories be similar for both samples.

There is another situation which does not occur often in practice, but which nevertheless is interesting from a theoretical standpoint. Assume that we wish to interpret a coefficient of reproducibility based on some other sample size than the usual 100. For some reason only 80 respondents are available. Can we ~~assume that 80 respondents will sup~~ply us with 80 per cent as much ~~information as 100 respondents in the~~ sense that the associated H values for ~~the two testing situations would be in~~ the ratio of 80 to 100? This assumption would not be justified. The processes of combining categories and selecting a subset of items were both selective processes, and thus a definite structuring of the situation took place favoring a reduction of the proportion of nonfitting responses. But there is no such selection in the case above, and there is no reason to suppose that the *proportion* of errors will be changed.[3] Since the degree of structuring is not changed, there will

[3] As R. L. Thorndike has pointed out to the author, in extreme reduction of the sample size there will be an appreciable reduction in the proportion of errors because of the fact that the responses among which we are looking for disagreement are also those which determine the scale values of the respondents. This spurious consistency is probably negligible, however, for values of N which are large in comparison to the number of possible scale values, which is equal to the number of items plus one.

~~is no difference in the homogeneity~~ value.

Thus it seems reasonable to assume that the mean of the distribution function of r is not appreciably affected by moderate changes in N, although the standard error of r obviously will be. When comparing two r's based on different sample sizes, we need not make any allowance for different entropy values; we need only keep in mind that the smaller the N, the larger the standard error of r. As present theory does not enable us to write the expression for this standard error, we can be safe only by avoiding sample sizes below the accepted standard of 100.

Table 1 gives the entropy values associated with dichotomous items for various differences between the response frequency proportions.

TABLE 1

ENTROPY VALUES ASSOCIATED WITH DICHOTOMOUS ITEMS FOR VARIOUS DEGREES OF RESPONSE BIAS

$p_1 - p_2$	Item Entropy
.00	1.0000
.10	.9928
.20	.9709
.30	.9341
.40	.8823
.50	.8113
.60	.7219
.70	.6098
.80	.4690
.90	.2864
1.00	.0000

More extensive tables of values for $-\Sigma p \cdot \log_2 p$ have been published (8), or if necessary, a table of common logarithms may be used to compute $\log_2 p$ values by use of the following relationship:

$$\log_2 p = \log_{10} p / \log_{10} 2 = 3.3219 \cdot \log_{10} p.$$

REFERENCES

1. Edwards, A. L., & Kilpatrick, F. P. Scale analysis and the measurement of social attitudes. *Psychometrika*, 1948, 13, 99–114.

2. Festinger, L. The treatment of qualitative data by scale analysis. *Psychol. Bull.*, 1947, 44, 149–161.

3. Goodenough, W. A technique for scale analysis. *Educ. psychol. Measmt*, 1944, 4, 179–190.

4. Guttman, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 80, 139–150.

5. Guttman, L. The Cornell technique for scale and intensity analysis. *Educ. psychol. Measmt*, 1947, 7, 247–280.

6. Loevinger, Jane. The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychol. Bull.*, 1948, 45, 507–529.

7. Mosier, C. I., & Frank, J. C. Statistical indeterminacy and uniqueness of the measure. *Psychol. Rev.*, 1947, 56, ...

8. Newman, E. B. Computational methods useful in analyzing series of ... data. *Amer. J. Psychol.*, 1939, 64, ...

9. Shannon, C., & Weaver, W. *The mathematical theory of communication*. Urbana: Univer. of Illinois Press, 1949.

10. Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, Shirley A., & Clausen, J. A. *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950.

11. Wiener, N. *Cybernetics*. New York: Wiley, 1948.

NOTE ON SCORE TRANSFORMATION AND NONPARAMETRIC STATISTICS

WALTER C. STANLEY
Brown University

It is becoming increasingly popular to analyze the statistical significance of markedly skewed data (e.g., latency of running or bar-pressing responses in rats) by means of nonparametric statistics. This point was particularly viewed for marked ease of calculation. However, one source of

The smaller of the two sums of ranks of the signs is therefore 8, and this difference is not significant, a sum of 4 being required for a *p* of .05.

In the right half of Table 1 the same statistical test is carried out on the differences in the reciprocals of these scores. Again the rank is

TABLE 1
ILLUSTRATIVE DATA ANALYZED IN TWO WAYS WITH WILCOXON'S PAIRED REPLICATES TEST

Subject	Raw Scores				Reciprocals			
	Cond. A	Cond. B	Diff	Rank	Cond. A	Cond. B	Diff	Rank
A	1	3	2	2	1 00	33	67	8
B	2	8	6	6	50	12	38	7
C	3	7	4	4	33	14	19	5
D	4	9	5	5	25	11	14	4
E	5	6	1	1	20	17	03	2
F	25	16	-9	-8	04	06	-02	-1
G	2	5	3	3	50	20	30	6
H	7	15	8	7	14	07	07	3

labor saving—working directly with raw data rather than with transformed scores—may sometimes be a mixed blessing. This could be the case when working with ranks of differences between paired scores. Here some score transformations would alter the magnitudes of the differences and thus their ranks.

Consider Wilcoxon's (1) test for paired replicates applied to the raw scores in Table 1. These scores could be latencies in seconds of a running response in rats obtained under two conditions of extinction. The largest difference is in the negative direction.

assigned a negative sign, but here this rank is 1, and the *p* obtained is between .02 and .01.

Clearly some consideration must be given to the meaningfulness of scale units, and more generally, to the population of values to which one wishes to generalize in order to take full advantage of this "rapid approximate" statistical technique.

REFERENCE

1. WILCOXON, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.

Received December 26, 1953.

Osgood, Charles E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953. Pp. xii + 800. $10.00 ($14.00 trade edition).

According to the preface, this book was written "to provide undergraduate majors and graduate students in psychology with a text that evaluates experimental literature in close relation to critical theoretical issues." For such use it has several virtues. Not only are the theoretical issues about which the book is centered important in their own right, but they also provide schemata that should help the student to comprehend and retain the myriad facts of experimental research. Osgood's description of many experiments is so minute as to be the next best thing to the originals themselves. Furthermore, within its range, the book reflects with marvelous faithfulness the problems and atmosphere of experimental psychology as this is represented in the latest volumes of American psychological journals.

This timeliness is apparently not due to unusual recency in the citations. A sample of 100 was drawn from the 1,290 references in the bibliography; when these were distributed by publication date, their median date was found to be 1937. A similar sample of 100 from the approximately 1,800 references in Woodworth's *Experimental Psychology* of 1938 (with which Osgood's book is bound to be compared) yielded a median date of 1922. Thus, the median ages of the citations at the time of the publication of the two books were both 16 years. (Is there a law here?) Woodworth's

references are, however, more skewed toward early dates than Osgood's.

On the other hand, the book is both massive and difficult. Two or three pages on vacuum tube amplifiers and cathode ray oscillographs will be insufficient for the student who is untrained in physics and superfluous needed for one so trained. There is a bit of mathematics here and there, but probably not enough to frighten the bright and well-motivated students for whom the book must have been written. The first 300 pages presuppose rather extensive acquaintance with neural anatomy and neurophysiology; a few anatomical drawings, including maps of cortical cytoarchitecture, would have been helpful here.

Then, too, for all of its 727 pages of text, the book is limited in scope. The author says that he has covered " ... the major part of what is called experimental psychology, including sections on sensory processes, perception, learning, and symbolic processes." In 1938, Woodworth had chapters on these topics plus three on feeling and emotion and one each on "experimental esthetics," GSR, reaction time, attention, and reading. The lack of material on motivation and action, which are treated at least in part by such a recent book as Stevens' *Handbook*, as well as other topics that other readers will no doubt miss, will be felt keenly in any broadly conceived proseminar in which Osgood's book is used. It is also likely to raise the tiresome question, what *is* experimental psychology?

The word "method" appears in the title but not as a heading in the sub-

ject index, and it can be said that the text deals more with method in ... than with method in general. One extensive evaluation and comparison of methods as such does ... methods, that on ... ... that is not wholly satisfactory. The method of average error is described inadequately and, the reviewer believes, incorrectly. The "flicker-fusion method" is given treatment coordinate with that of the standard methods even though it must itself employ one of these latter methods. ... it appears that the author is more worried than he need be over the interpretation of the results of the two-category method of constant stimuli.

To regard this work merely as a textbook or handbook would, however, overlook its most significant aspirations, which have to do with the evaluation of theory and the maintenance, in the author's words, of "a certain continuity of approach." Now, it is notorious that experimental psychology does not easily submit to systematic unification, and we regretfully judge that this proposition is not threatened by the present book. The theoretical atmosphere of the first six chapters, on sensory and perceptual processes, is noticeably different from that of the remaining ten, on perceptual dynamics, learning, thinking, and language. Osgood's Hullian views have no relevance to such matters as the cortical basis for sensory quality and intensity, the quantal hypothesis in audition, and the significance of visual adaptation. What little integration is achieved for sensory psychophysics and psychophysiology comes from the author's interest in the mode of action of the cerebral cortex, where he prefers neurostatistical conceptions to the "dynamics" of Köhler. Hel-

... which ... together ... unconnected, is not mentioned.

... ter on central factors in perception ... remainder of the book, consideration is given to major theorists. Tolman, Guthrie, the gestalt psychologists, and Hull are ... and briefer accounts are given of Hebb, Skinner, and others. It is obvious that ... has made intense efforts to understand and to present sympathetically opinions with which he partly or wholly disagrees, as well as to exhibit in full relief the weaknesses of those by which he takes his own stand. How successful he has been in the exposition of each of these views will certainly be judged differently by readers of different theoretical predispositions. The reviewer's belief is that the systems of Hull, up through *Principles of Behavior*—the later books were too recent to be included), Guthrie, and Tolman are discussed in such a way as to give their followers little cause for complaint; indeed, Osgood has constructed challenging formalizations of the last two, with the aid of Voeks in Guthrie's case. Gestalt theory is more alien to him, and he makes the mistake of writing as if Köhler, Wertheimer, Koffka, Lewin, and J. F. Brown adhered to one system in common.

Three pages are allotted to Hebb as an exponent of physiological theory. Unfortunately, the more purely psychological aspects of Hebb's ideas are lost sight of, so that, for example, his contributions to the theory of thinking are not referred to in Osgood's chapters on this topic. Least satisfactory is the treatment accorded to Skinner. His views on

the role of theory in psychology are omitted, and some of the remarkable regularities that he has established, which are clearly of concern to a learning theorist, are given little or no attention. This is true, for example, of the effects of different reinforcement schedules. Periodic reinforcement is mentioned in passing, with the comment that "The mechanism whereby an animal smoothly adjusts its rate of response under these conditions is unknown." "Intermittent reinforcement" is discussed more thoroughly, but Skinner's name is not mentioned in connection with it. The same is true of the problem of the number of kinds of learning. A rather misleading statement (p. 307) could cause the reader to suppose that Skinner accepts the concept of disinhibition.

Osgood's own "mediation hypothesis," by which he hopes to effect a rapprochement between "reinforcement" and "cognitive" theories, is not a hypothesis in the ordinary sense, but rather a broad extension of Hull's concepts of anticipatory goal responses and their associated proprioceptive stimuli. The following statement from the chapter on language behavior exemplifies one of the functions attributed to mediators, that of serving as *signs*:

... a pattern of stimulation which is not the object is a sign of the object if it evokes in the organism a mediating reaction, this (a) being some fractional part of the total behavior elicited by the object and (b) producing distinctive self-stimulation which mediates responses which would not occur without the previous association of nonobject and object patterns of stimulation (p. 696).

Such mediated self-stimulation has cue, motivating, and reinforcing properties. Mediators are not necessarily peripheral but may be purely cortical events; they are said to have the status of "hypothetical constructs." The following comments

are made by the reviewer, with the caveat that they should be taken as no more than fragments of a possible critical analysis of the hypothesis. (a) The mediators are so substitutable for the "ideas" of association theory that occasionally the discussion sounds like a more translation of such theory into Hullian language. It would be no defense to attack cognitive theory on the same grounds. (b) While the evidence submitted in favor of the hypothesis seems to indicate that some sort of mediation is involved in the behaviors that are described, it does not, in the reviewer's opinion, discriminate clearly in favor of fractional goal responses. (c) The explanations by mediators have an essentially chainlike character, and are open to all of the objections that have been raised against such a conception since Dewey's paper on the reflex arc. (d) The attribution to mediators of both motivating and reinforcing properties leads to the same extreme difficulties that leave the stimulus-reduction theory of reinforcement hanging by a thread on page 443. (e) One would like more evidence that the hypothesis can *predict* new and unexpected phenomena.

On questions of philosophy of science and theory of knowledge the book is disconcertingly artless, and the thought is more energetic than subtle. Although Osgood calls himself a materialist, and says that "Behaviorally, . . . the environment is a pattern of neural energies in the central nervous system," and that, when one touches something, "The awareness of sensation is not . . . in the fingertips . . . but in the brain," he nevertheless wrestles with the ghosts of introspectionism without vanquishing them, since they remain to haunt many pages of the book. At the same time, he shows no great sensitivity to the problems of earlier days;

e.g., almost anything that one might say about oneself is called "introspection." ... For examples from other contexts, Lloyd Morgan's canon is treated as if it were subject to experimental tests, and the discussion of regression leads to the methodological, awkward conclusion that motivated forgetting " ... is certainly a valid observation in the clinic and probably would be verified in the laboratory. ... "

So ambitious a work can be expected to contain some inaccurate and debatable matter. Thus, the all-or-none law is so stated as to make it depend upon the full utilization of "materials" in the neuron; in connection with the Talbot-Plateau law, it is asserted inaccurately that " ... with a light-dark ratio of one to one, the fused field will appear one-half as bright as the illuminated sector"; the color solid, which some might regard as a minor triumph of taxonomy, is dismissed as a mere pedagogical device, apparently on the incorrect supposition that it was intended to represent the facts of color mixture; it is said that interference theories of extinction never specify the interfering responses, whereas a number of such responses were identified two pages earlier in the description of research by Wendt; a reference to work by Birenbaum and Zeigarnik leaves the impression that Lewinian theory presumes that boundaries between regions within the person are less permeable in the child than in the adult, when in fact the theory asserts the reverse.

The book is seriously marred by numerous errors of grammar, spelling, and typography, as well as other blemishes of which some are matters of taste. It is regrettable that most of these were not weeded out by the publisher's editing. A few miscellaneous examples follow.

The names Bartlett, Brown, Lloyd, Hovland are sometimes misspelled; a careful reader, on the point of giving up, finds on a certain line, the recurring hint of a wan similar mistake in the abscissa of Fig. 24, ... seems to be a printed, etc. To typographical errors perhaps most troublesome is the printing of a negative exponent in an exponential expression as a subscript (p. 16). The more common error, captions are generally too sparing, and the text does not always refer to, or even accurately to the figures. Thus, the text refers to dashed and solid lines in Fig. 12, but only the latter are found; Fig. 5 is too sketchy for the text; hair cells are mentioned in the text but not identified in Fig. 5 and Fig. 24, the abscissa of Fig. 41 should be labeled ... seem labels are needed for Fig. 207; the text refers to Fig. 117 when Fig. 115 is intended. (On the other hand, some of the figures are excellent, as, e.g., Fig. 89 and 101.) It may also be mentioned that the type seems very small for the length of line on the large single-column page.

The foregoing, however distressing, are minutiae, while Osgood's accomplishments are not. Scattered throughout the book are original theoretical and experimental contributions on topics ranging from color contrast to transfer and retroaction in learning. Several chapters are unusually interesting, e.g., Chapter 7 on perception and Chapter 14 on problem solving. Furthermore, the author is at his best in assembling the evidence for and against testable research hypotheses and doggedly following the trail of the experiments even when they lead in an undesired direction; and this is, after all, his principal intention.

FRANCIS W. IRWIN.
University of Pennsylvania.

HAVIGHURST, ROBERT, J., & AL-BRECHT, RUTH. *Older people.* New York: Longmans, Green, 1953. Pp. xvi+415. $5.00.

This book is an admirable antidote to the prophets of doom and gloom who identify aging with senility, disa-

bility, and frustration. It is based on a detailed survey of a small midwestern city (7,000 in the town itself and 4,000 in the adjacent trading area) and presents factual data about what the older people there are like (670 of them 65 or older). Individual interviews were carried out with a sample of 100 of the old people drawn to represent the entire old age population by sex, socioeconomic status, and marital status.

The book, however, is more than a compilation of factual information. It adds a conceptual framework for the analysis of the needs and desires of individual aging people, an analysis of the community's reaction to the aged, and a formulation of basic principles essential to the effective planning of programs for the aged. Writing in an easy, flowing style, with enough ancecdotal material to maintain a lively interest, the authors manage to impart a great deal of factual information without the reader's being aware of it.

The book should be general in its appeal, not only to those who are especially concerned with the problems of aging, but also to the aging people themselves. Thus, the physician who may be exposed daily to the aches and pains of older people will be agreeably surprised to learn that 79 per cent of the oldsters in this cultural environment regarded themselves as "healthy" and that only 6 per cent were homebound and 2 per cent actually bedridden. The social worker, who must deal with the problem of finances, living arrangements, etc. among the aged, ought to know that 43 per cent of this population reported they were happily situated and that only one-fourth were actually unhappy. It is also important to know that happiness was not related to economic status and that health

became a major component of happiness only in the 20 per cent of the population who had specific complaints; i.e., old age was what the individual made it, with economic and health aspects assuming a secondary role. All but 4 per cent of the population reported that they had enough money to get along on provided they had no unusual medical expenses. Anxiety about what they would do in case of protracted illness requiring hospitalization was present in a large proportion of the elderly people. There was no evidence that the community rejected old people because of their age. However, the importance of the role of the individual in the community was emphasized. In general, the community seemed quite permissive about what it expects of older people and looked with favor on continued activity.

The authors present a number of concepts that may require further analysis. For example, their outline of the meaning of work shows the need for a variety of retirement plans if we are to meet the needs both of the individual and of society. The concept of socioeconomic mobility is applied to the psychological adjustments of aging people in an interesting manner. Whether all of the conceptual formulations will prove useful remains to be seen, but the authors are to be complimented on their willingness to theorize and thus reduce a mass of specific observations to some intelligible formulation.

Psychologists with a quantitative psychometric viewpoint will be disappointed by the use of nonscaled questionnaire techniques, and sociologists may be disturbed by the inclusion of a chapter on "A Personal and Social Philosophy of Old Age" which makes specific recommendations about "rational defenses" that can be

used by older people. One might also quibble about whether the population of Prairie City is representative of the entire United States. However, the solid facts of how one group of aging people behaved in a definable environment have been admirably explored by this study. Only by similar studies in different community environments can the generality of the findings be settled.

It is good to know that in some community structures a large proportion of the older people can optimistically meet the problems of aging. This is a book that should be read by everyone interested in the welfare of his own community.

<div align="right">N. W. SHOCK.</div>

Baltimore City Hospitals.

FEDERN, PAUL. *Ego psychology and the psychoses.* (E. Weiss, Ed.) New York: Basic Books, 1953. Pp. 375. $6.00.

The editor had the difficult task of organizing the presentation of a systematic theory of ego functioning out of another man's half-century (1901–1952) of prolific publishing (almost 100 articles, lectures, etc.). The 40-year association as pupil and friend of Paul Federn provided important background for this task. In the 21-page introduction, Weiss attempts to establish Paul Federn's orthodox loyalty to Dr. Freud (to take care of the "minor" differences in the area of ego theory), to apologize for Federn's difficulties in exposition which is " . . . very rich in content but . . . often complex etc." (p. 21), and to present a condensed, clarifying guide to the metapsychological wilderness of the text. He apparently did not feel that Federn could provide in 16 articles and 340 pages a sufficiently clear presentation of his thoughts, but unfortunately the guide itself needs a

guide. The first five "theoretical papers" are a maze of unclear verbal gymnastics, poorly defined concepts, and repeated self-adulation by Federn which does not sufficiently reward the reader for his difficult labor of seeking meaning. In this part of the book, Federn shares a difficulty common to many psychoanalytic authors who consider operationalism a primitive scientific procedure and prefer an artistic verbal medium.

The second part of the book (nine papers) is a fuller measure of Dr. Federn's contribution. Here, his rich clinical experience and observational astuteness appear to good advantage. A conception of the ego differing in many respects from that postulated by Freud is presented. "Ego Feeling" seems to be viewed as a basic energy which integrates ego functioning and makes ego experience possible. An interesting hypothesis suggesting a somatic and mental differentiation of ego feeling is developed. There is a much-repeated emphasis on psychosis as an ego-defeat phenomenon, as against the neurotic dynamics of ego defense. The so-called "Steinach effect" is defended as a sufficient basis for the sterilization of latent psychotics and psychotics. However, the psychological rationale given in Federn's papers hardly constitutes a convincing argument.

This book is not the "Vade Medicum" for the treatment of psychosis which the publishers enthusiastically proclaim on the book-jacket blurb. It is certainly not recommended for general reading. However, if the reader is familiar with psychoanalytic terminology and has the perseverance to follow Federn through "thick and thick," then he may find in the papers many provocative speculations based on extensive clinical experience. There might even be the reward of

deriving fruitful hypotheses for experimental investigation from some of these speculations.

M. ERIK WRIGHT.
University of Kansas.

McCURDY, HAROLD GRIER. *The personality of Shakespeare.* New Haven: Yale Univer. Press, 1953. Pp. xi+243. $5.00.

Almost twenty years ago Caroline Spurgeon published the results of her extensive study of Shakespeare's figures of speech from which she drew a number of inferences concerning Shakespeare's interests, preoccupations, attachments, personality traits, personality development, and the like.[1] Her study was based upon the premise that a person reveals things about his personality by the kinds of metaphors he uses.

McCurdy has tried to get at the personality of Shakespeare by a different route. He has counted the number of lines that every character in every play speaks in order to determine the principal characters in each play. The assumption is made that the more important features of Shakespeare's personality are represented in the principal characters. In addition to character analysis, McCurdy has analyzed some of the major themes that run through a number of plays much as one would analyze a TAT protocol.

For anyone who admires Shakespeare and who feels that the psychologist has something to offer in the way of method for shedding more light upon the kind of person that Shakespeare must have been in order to write as he did, this is a fascinating study. Dr. McCurdy is dedicated to his subject and writes with sensitivity and critical devotion.

Most psychologists do not quarrel

[1] *Shakespeare's imagery and what it tells us.* New York: Macmillan, 1935.

with the hypothesis that a writer projects himself into his writings and that it should be possible therefore to make a personality analysis of a writer from his writings alone. The same hypothesis probably holds for painters, composers, architects, designers, and anyone who produces something out of his imagination. However, Shakespeare is the worst possible subject on whom to test the hypothesis. Virtually nothing is known about the man. In fact, so little is known that a number of people believe that someone else must have written the plays attributed to Shakespeare. Such being the case, how can one ever hope to confirm or infirm the inferences that are made about the personality of Shakespeare from his writings?

I wish that Dr. McCurdy had devoted his considerable talents both as a psychologist and as a literary critic to the examination of a writer about whom a great deal is known. Hemingway or Mickey Spillane would make excellent choices. Is it true, for instance, that Mike Hammer represents the shadow-side of Spillane's personality? How much of the old man of the sea is to be found in Hemingway? McCurdy's methods of of analysis would undoubtedly strike pay dirt and reveal a great deal about the nature of projection if he used them on more appropriate subjects.

CALVIN S. HALL.
Western Reserve University.

STOLUROW, LAWRENCE M. (Ed.) *Readings in learning.* New York: Prentice-Hall, 1953. Pp. viii+555. $6.00.

Stolurow has collected 42 bibliographical items from the field of learning, either in whole or in part, and by combining three into one unit and two into another, has presented 39 articles under one cover. His pur-

pose was to provide a sample of original material to supplement secondary source materials since "only by reading and analysing original reports can the student learn how to conduct a variety of different types of research and at the same time become aware of the problems and difficulties involved." Presumably this book was intended to serve as a second text in a course in theories of learning or perhaps as a source book in a course in which no text was used, although Stolurow suggests that, "Where a laboratory is available, this volume could serve as a laboratory text. The experimental studies could be used as models for both experiments and report writing, and the theoretical articles as bases for new studies."

It is a fairly safe estimate that at least 5,000 bibliographical items have been published in the past 70 years from which Stolurow chose 42. His selections could well be analyzed in terms of whether they are essentially systematic as opposed to experimental articles, in terms of how well they represent the various problem areas that have been of concern to people interested in learning, in terms of how well various theoretical points of view are represented, and in terms of whether these particular items are in fact good models.

The book is divided into eight sections or chapters. The first of these, titled "Some Systematic Positions," contains six units representing eight of the 42 articles selected. These eight and seven others in later chapters, something over 35 per cent of the selections, are essentially nonexperimental. In addition, "Every chapter and article contains some theory. This is a sign of the times." Thus the book is very heavily weighted in favor of theory and its exposition, although it does contain 27 detailed research reports.

Stolurow's selections are representative of points of view. He reports that in himself his conviction as five points of quality and the additional criteria that selections should be recent rather than old, human rather than animal, and S-R rather than gestalt. His character or systematic positions has selections from Thorndike, Hull, Guthrie, Estes, and Skinner, representing S-R positions, and Tolman, who is somewhat difficult to classify. The other systematic or nonexperimental articles are by Hull, Mowrer, Pavlov, Skinner, Spence, Tolman, and Woodworth.

Aside from the S-R bias, which is confessed and deliberate, the nature of the selection process is partially revealed by who is omitted. There are, by rough count, 947 references cited in this list of 39 items. Thirteen men have ten or more references on this list. Stolurow has selected articles from eight of these men, but has not selected any from the writings of Hilgard, Krech (Krechevsky in the bibliography), McGeoch, Maier, or Neal Miller. Each has published at least one "classic" article in the period from which these items were selected. Twenty-two were published between 1947 and 1950 and the remaining 20 between 1928 and 1946.

The 27 articles which report experiments in detail are heavily weighted in favor of experiments involving human subjects (18) as against those involving animals (9). This weighting follows largely from the choices of subject matter represented. There are two experimental articles on conditioning concepts and techniques, four on motivation and reinforcement, six on motor and verbal learning, three on discrimination and perceptual learning, two on educational and social learning, five on retention and forgetting, and five on transfer.

As models of research design, there can be little quarrel with Stolurow's selections. They are some of the better examples of the kind of article usually found in the *Psychological Review*, the *Journal of Experimental Psychology*, and the *Journal of Comparative and Physiological Psychology*, from which most of his selections were taken. They are also typical of the writing style of these journals and are good models if one wishes to perpetuate this style. If one wishes to teach a more intelligible style, some of these selections can be used as examples of what not to do.

The book is printed by a photo-offset process. The right-hand margins of the pages are not justified. The paper used is sufficiently transparent that printing from the back of the page and from the next page shows through. The binding is a hard cover, but it is covered with paper rather than cloth and the spine of my copy is broken on both sides from the handling it has received in preparing this review. It is a shabby job of book making and, from this standpoint, overpriced by a very wide margin.

EDWARD L. WALKER.
University of Michigan.

PSYCHOTHERAPY RESEARCH GROUP, PENNSYLVANIA STATE COLLEGE, (Wm. U. Snyder, Chairman). *Group report of a program of research in psychotherapy.* State College: Pennsylvania State Coll. Press, 1953. Pp. iii+179. $2.25.

Regardless of divergent viewpoints about the merits of client-centered counseling, much credit is due Rogers and his students for their pioneering and ingenious research with verbatim interview recordings. This report is part of a sequence which began little more than a decade ago, but it already represents a second generation effort, being the collaborative production of nine doctoral students directed by William U. Snyder. The report consists of a description of the research program, nine thesis condensations, a summary and discussion by Dr. Snyder, a bibliography, and a detailed appendix that provides the rating and coding procedures utilized in the studies.

The basic sample of cases studied was 100 student counselees from which maximum N samples were drawn depending upon various selection criteria, e.g., number of interviews, transcribability of interviews, tests taken, etc. In addition to analyses based upon the recorded interviews, use was made of pre- and postcounseling tests. These tests were the Rorschach, the MMPI, and the Mooney Problem Check List. Further data were provided by ratings made by clients, counselors, and independent judges using such devices as a post-counseling client scale, a therapist personality scale, and a posttherapy counselor check list. Somewhat surprisingly in view of recent trends, no Q sorts were employed.

Among the topics investigated were the reasons for early dropouts, the relationships between counselor and client characteristics, the development of a composite criterion for measuring client progress, the predictability of client verbal behavior during counseling, indices of resistance, and comparison of the characteristics of more and less successful cases.

A number of new hypotheses and variables have been introduced in these studies and, despite the fact that most of the findings are inconclusive, the report is an important contribution to research methodology. Many readers will doubtless be disturbed by the continuing defense offered for the position that the effec-

uccess of therapy can be evaluated
without the necessity of external cri-
teria.

LEONARD S. KOGAN.
Institute of Welfare Research,
Community Service Society of
New York.

STEPHENSON, W. *The study of behav-*
ior. Chicago: Univer. of Chicago
Press, 1953. Pp. ix+376. $7.50.

If Stephenson had set forth his
purposes in writing this book, which
carries the subtitle "Q-technique and
its methodology," the reviewer's task
would be easier. Our inference is
that his aims were the following: (a)
to challenge much of current method-
ology in psychology, (b) to explain
Q-methodology and (c) to show by
illustration how it can put psychol-
ogy's "house in scientific order," and
(d) to demonstrate that theory test-
ing and scientific conclusions are pos-
sible on the basis of a single case.
Perhaps when the author speaks of
the "platform upon which we are to
campaign," he is telling us that his
sole aim is to promote Q-methodol-
ogy.

We shall not attempt to list here
all of the concepts and all of the
people against which and whom
Stephenson arrays himself for battle.
He does not have any faith in ordi-
nary factor analysis (R-methodol-
ogy), in measurement, in norms, in
large samples, or in any so-called
generalizations springing therefrom.
He admits that he alone is "in step
and all others out" (p. 348), but this
does not keep him from citing what-
ever supporting fragments he can
find, whether these be found in the
writings of J. R. Kantor or of J. M.
Keynes or of some very obscure per-
son. His sallies, courageously set
forth, will be found either interesting
or irritating, according to the pro-
clivity of the reader.

With regard to purpose b one would
expect that an author who confesses
because such intellects as Godfrey
Thomson and Cyril Burt have not
understood his writing would make a
special effort at clear and concise
exposition. Instead, we find a poorly
organized, piecemeal presentation,
more confusing than enlightening.
Thus, we can anticipate continuing
misunderstanding, and consequent
misuse of Q-methodology.

The second half of the book is
devoted mainly to applications of Q-
methodology in the areas of type psy-
chology, questionnaire analysis, so-
cial psychology, self-psychology, per-
sonality, projective tests, and clinical
psychology (a chapter to each area).
We are told that "Q-technique has
its applications in almost every nook
and cranny of psychology in its re-
search aspects" (p. 338) and "in every
branch of psychology where *behavior*
is at issue" (p. 343). If we are to
judge from the given illustrative ap-
plications, the quoted claims repre-
sent wishful thinking.

Clinicians and others who, by
necessity or by choice, deal with a
single case will find comforting re-
assurance and be motivated to read
further by "We are to work, instead,
with a single person, at the call of a
theory. Yet we shall reach valid,
scientific conclusions" (p. 5). The
continuous stress on the merits of
the single case leads ultimately to "In
principle, one may work scientifically
for a lifetime with a single case" (p.
343). Unfortunately, by the time
one has spent a lifetime developing a
set of principles for predicting (or
explaining) every fragment of be-
havior of a single case, the subject
will have ceased to behave. Or
another logical conclusion to this sort
of thing is that psychologists must
develop two and a half billion "sci-
ences" to explain the behavior of the

two and a half billion human inhabitants on this planet! What of the task of animal psychologists?

The fact that our comments have been restricted to general reactions should not be misconstrued as indicating that the margins of the reviewer's copy of the book are free of specific questions. Far from it.

QUINN McNEMAR.

Stanford University.

GOLDHAMER, HERBERT, & MARSHALL ANDREW W. *Psychosis and civilization*. Glencoe, Ill.: The Free Press, 1953. Pp. 126. $4.00.

This slim volume consists of a reprint of two statistical studies in the frequency of mental disease. The first and more significant investigation is concerned with an analysis of admissions to mental hospitals in Massachusetts and Oneida County, New York, extending back to 1840. Consistent with data previously reported by others for the past half century, the present authors found that for age groups under 50 there has been no increase in the frequency of psychoses over the past 100 years. This finding should put an end to the recurring myth that psychoses are a product of the stress and strain of modern life. The second paper presents expectancy rates of mental disease. It differs from earlier studies in that the tables prepared state the risk of admission to a mental hospital between any two points of an individual's life. This method of presentation serves to accentuate the high incidence of admissions for the older age group. If a 60-year-old male survives to age 85 he runs a 10 per cent risk of being admitted to a mental hospital.

JAMES D. PAGE.

Temple University.

FRENCH, THOMAS M. *The integration of behavior*. Vol. II. *The integrative process in dreams*. Chicago: Univer. of Chicago Press, 1954. Pp. xi +367. $6.50.

When the history of twentieth century psychology is written, it will undoubtedly be characterized as the period when two great streams of psychological thought, one flowing from the laboratory, the other from the clinic, converged and ran together to form a unified science of dynamic psychology. The future historian will observe that the dialectical process which eventuated in the synthesis of experimental and clinical psychology took a long time and had to overcome many obstacles. Among the obstacles that he will discuss, one at least is apparent to us today, namely the *insularity* of the proponents of each of these major orientations. This insularity prevents one side from communicating with the other, so that each remains ignorant of what the other stands for.

Fortunately for psychology there are indications that the iron curtain of insularity is being lifted and that a real exchange and integration of ideas are beginning to take place. The name-calling era is drawing to a close. Psychologists brought up in the tradition of experimental psychology are reading and being inseminated by psychoanalysis, and psychoanalysts, though to a lesser extent, are reading and being inseminated by experimental psychology. An outstanding example of a psychoanalytically trained investigator whose thinking has been fertilized by intercourse with systematic experimental psychology is Thomas French, associate director of the Chicago Institute of Psychoanalysis.

In a number of articles published during the last 20 years, French has

demonstrated his ability to synthesize the two orientations. Now, in his impressive work in progress, three volumes of which are still to be published, we are privileged to witness the culmination of his integrative endeavors. The present volume is an application of the basic postulates set forth in Volume I to an understanding of dreams.

Essentially, what French has done is to graft Tolman's cognitive theory onto Freud's motivational theory. French's key concept is *cognitive structure*, by which he means "a hierarchy of plans for achieving an end-goal." Implicit in this definition is the concept of motive since every plan involves striving toward a goal.

French not only believes that every dream has a cognitive structure but also that the cognitive structures of different dreams of the same person form a pattern. The cognitive structure of a dream is discovered by using various sources of evidence, namely, information about the dreamer, free associations, translation of symbols by functional analysis, and comparisons of one dream with the other dreams of a series. All of this information is blended together by employing the method of internal consistency, which is the favorite method of psychoanalytic investigators. The end result is a comprehensive understanding of the dreamer's conflicts, their relative intensities and inter-relationships, their historical origins and contemporary significance, their underlying physiological and motivational patterns, the dreamer's plans and capacities for resolving them, and his hopes of success.

The rhetorical strategy of the book consists of discussing the dreams of a patient who is undergoing the type of psychoanalytic treatment practiced at the Chicago Institute of Psychoanalysis. Some readers may question whether the strategy is a successful one. This reader found that it became quite tedious to try to follow and bear in mind all of the intricacies of analyzing the dream series. It may well be that every idea is related in some way to every other idea, and that there is almost an infinity of components and levels in the cognitive structure of a particular human mind, but I wonder whether there is not a better way to get this across. Perhaps not. It seemed to me that there was an unconscionable amount of redundancy and that more severe editing could have made the book more readable. French's use of diagrams is an aid to quick understanding and should have enabled him to abbreviate the extended discussions.

In spite of these literary defects the book is a solid contribution to the major task of twentieth century psychology.

CALVIN S. HALL.
Western Reserve University.

ANASTASI, ANNE. *Psychological testing.* New York: Macmillan, 1954. Pp. v+682. $6.75.

AUSUBEL, DAVID P. *Theory and problems of adolescent development.* New York: Grune & Stratton, 1954. Pp. xviii+580. $10.00.

BELLOWS, ROGER M., & ESTEP, M. FRANCES. *Employment psychology: the interview.* New York: Rinehart, 1954. Pp. xxi+295. $4.25.

BRAY, DOUGLAS W. *Issues in the study of talent.* New York: King's Crown Press, 1954. Pp. xi+65. $2.00.

CARMICHAEL, L. (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1946, 1954. Pp. ix+1295. $12.00.

CHINEY, ELY. *Sociological perspective; basic concepts and their application.* Garden City: Doubleday, 1954. Pp. vi+58. $.85.

CRONBACH, LEE J. *Educational psychology.* New York: Harcourt, Brace, 1954. Pp. viii+628. $5.50.

FINK, KENNETH H. *Mind and performance; a comparative study of learning in mammals, birds and reptiles.* New York: Vantage, 1954. Pp. xi+113. $3.00.

FINLAY, WILLIAM W., SARTAIN, A. Q., & TATE, WILLIS M. *Human behavior in industry.* New York: McGraw-Hill, 1954. Pp. xi+247. $4.00.

FRUCHTER, BENJAMIN. *Introduction to factor analysis.* New York: D. Van Nostrand, 1954. Pp. v+280. $5.00.

DE FOREST, IZETTE. *The leaven of love.* New York: Harper, 1954. Pp. xvi+206. $3.50.

FREUD, SIGMUND. *The origins of psycho-analysis; letters to Wilhelm Fliess, drafts and notes: 1887–1902.* (Marie Bonaparte, Anna Freud, and Ernst Kris, Eds.) New York: Basic Books, 1954. Pp. xi+486. $6.75.

FRYER, DOUGLAS H., et al. *General psychology.* (4th Ed.) New York: Barnes & Noble, 1954. Pp. xix+300. $1.50.

GEDDES, DONALD PORTER. (Ed.) *An analysis of the Kinsey reports on sexual behavior in the human male and female.* New York: E. P. Dutton, 1954. Pp. x+319. $3.50.

GINZBERG, ELI, et al. *Psychiatry and military manpower policy.* New York: King's Crown Press, 1953. Pp. xi+66. $2.00.

JERSILD, ARTHUR T. *Child psychology.* (4th Ed.) New York: Prentice-Hall, 1954. Pp. v+676.

KATZ, DAVID, CARTWRIGHT, DORWIN, ELDERSVELD, SAMUEL, & LEE, ALFRED MCCLUNG. *Public opinion and propaganda.* New York: Dryden, 1954. Pp. xx+779. $6.25.

MCGILL, V. J. *Emotions and reason.* Springfield: Charles C Thomas, 1954. Pp. xiii+122. $3.25.

MICHAL-SMITH, H. (Ed.) *Pediatric problems in clinical practice; special medical and psychological aspects.* New York: Grune & Stratton, 1954. Pp. x+310. $5.50.

PENNINGTON, L. A., & BERG, I. A. *An introduction to clinical psychology.* (2nd Ed.) New York: Ronald, 1954. Pp. viii+709. $6.50.

REMMERS, H. H., RYDEN, E. R., & MORGAN, C. L. *Introduction to educational psychology.* New York: Harper, 1954. Pp. ix+435. $4.00.

RUNES, DAGOBERT D. *Letters to my daughter.* New York: Philosophical Library, 1954. Pp. 131. $2.50.

WULFECK, J. W., & BENNETT, E. M. *The language of dynamic psychology; as related to motivation research.* New York: McGraw-Hill, 1954. Pp. 111. $4.00.

ZILBOORG, G. *The psychology of the criminal act and punishment.* New York: Harcourt, Brace, 1954. Pp. xi+141. $3.50.

Psychological Bulletin

RESEARCH ON SENSORY INTERACTION IN THE SOVIET UNION[1]

IVAN D. LONDON
Brooklyn College

Since the early thirties the general area of sensory interaction has been the object of considerable attention by researchers in the Soviet Union.[2] However, owing to the language barrier and the near unavailability of the literature, most of the experimental and theoretical work that has been done in this field has gone unnoticed. Perhaps another factor tending to keep western attention averted has been the discouragement induced by the difficulty encountered in attempting to duplicate the reported findings.

It is true that much of the Soviet work on sensory interaction adheres to standards of execution, reportage, and interpretation that would be quite unacceptable to the western researcher. As a matter of fact, even a casual survey of the Soviet literature yields ready evidence of inadequate instrumentation and methodology, scanty detail, and a primitiveness in the statistical treatment of data which makes anything beyond an arithmetic mean a rare encounter.[3]

Nevertheless, western work on sensory interaction has been, in the main, scattered and desultory, whereas in the Soviet Union the subject has been given systematic and sustained attention (172, p. 77). For this reason alone a survey of the claimed experimental findings and of the correlative theorization seems to be in order.

The writer proposes first to outline representative basic findings, then to discuss the conditions that are said to affect interaction among the sense organs, and finally to sketch the theory that has been developed to account for this interaction. By

[2] Some western students of Soviet psychology have seen in this the hand of Marxist ideology (13, pp. 161–162). The real reasons for the concentration of research in this area, however, are probably more on the pedestrian side. The writer discusses them in a paper to be published elsewhere.

[3] The published reports and resolutions of various professional meetings (not to be confused with the stage-managed affairs of the Lysenko variety), along with the abstracts of the discussions following the reading of papers, provide a frank admission of some of these deficiencies (5; 15, p. 805; 59; 109; 116–118; 172, p. 55; 198, p. 387; 347; 382; 484; 485; 486, p. 249; 487, p. 157; 488, p. 623; 489, pp. 650–651; 492, p. 715; 495, p. 330). Recently standard deviations and coefficients of correlation have been making a spotty reappearance after their exit from the psychological scene in 1936. Tests of significance are still only occasionally encountered (13; 88; 197, p. 45; 257, pp. 79–80; 335; 339, p. 402).

research in sensory interaction[4] will be meant generally investigations that explore modifications of response in one sense organ under direct stimulation when another sense organ has been or is subject to its own characteristic stimulus. The latter stimulus is referred to in the Soviet literature as "indirect" (*nepriamyĭ*) or "accessory" (*pobochnyĭ*)—the last being the preferred term in this paper. Excluded for the most part from the survey will be intersensory research on (*a*) sense organs in pathological states (12, 75, 77, 153, 196, 475–477, 479); (*b*) nonsensory reactions (electrical and ionic excepted) of the sense organs;[5] (*c*) binaural interaction (1, 2, 197); (*d*) binocular and other intramodal interaction (pertinent exceptions being made, however);[6] (*e*) the effect of pain, as the accessory stimulus, on sensory response (375, 393, 474); and (*f*) the influence of interoceptive stimulation and dermal irradiation (62, 63, 135, 162, 179, 185, 187, 191, 314, 315, 317, 321, 355). The writer wishes it to be understood that all statements of ostensible fact in the paper refer to data or interpretation of data generally accepted as valid in the Soviet literature. The survey covers the period from 1930 to the present.

CHANGES IN SENSORY RESPONSE UNDER ACCESSORY STIMULATION

Absolute Sensitivity of Peripheral Vision

The absolute sensitivity of peripheral vision was early shown to under-go modification with excitation of the nonvisual organs (34). Modifiability of peripheral sensitivity was demonstrated also with excitation of the central visual system which, together with the peripheral system, constitutes the visual apparatus. Thus, illumination of the macular field was found to reduce sensitivity of the peripherally placed rods shortly after cessation of the stimulus (68, 73, 75, 93, 172, 286, 294, 326, 402, 403).

The antagonistic relationship surmised to obtain between the central and peripheral retina has been the subject of considerable study with results confirming and extending earlier findings (325, 329, 357–359, 403, 480). For example, it was noted that illumination of the central retina has no effect on the retinal periphery in the color blind, whereas, in those whose cone apparatus is nondefective, characteristic modifications of peripheral sensitivity are to be observed (265, p. 30; 357; 358). Thus, the reciprocal nature of the cone-rod relationship, which Orbeli had earlier hypothesized, came to be an accepted fact (69; 233, p. 473; 256, p. 47; 285; 371; 436–438).

While macular illumination, after its cessation, induces a reduction in peripheral sensitivity, a subsequent increase was found. Thus, after macular illumination with moderately bright light for one to ten minutes, a decrease in peripheral sensitivity for the first few minutes is exhibited, followed, however, by a period of supernormality (286–288, 400, 401).[7]

---

[4] Literally, interaction of the sense organs (*vzaimodeĭstvie organov chuvstv*).

[5] E.g., modifications of intraocular pressure, angioscotomata, etc. (384–388, 475–478).

[6] E.g., binocular contrast, various processes of retinal induction and sensibilization, simultaneous olfactory stimulations, etc. (3; 43–47; 48, pp. 209–215; 49; 53; 54; 111; 113; 128; 194; 195; 274, pp. 258–259; 451; 481).

[7] The first phase is considered as an instance of inhibition; the second, as one of disinhibition. However, "sensibilization" is not considered to be derived always from disinhibitory processes. Thus, the sensitizing action of threshold stimuli upon peripheral vision is ascribed to the increase of neural excitability caused by weak excitation of the visual apparatus (274, pp. 129–130, 185–186).

Not only does macular illumination of one eye effect a reduction in the sensitivity of its retinal periphery, but a diminution is to be observed likewise in the companion eye. Further study, however, revealed that more or less prolonged adaptation of one eye to moderately bright light results in a heightened peripheral sensitivity in the other, while increased central sensitivity is in turn reported subsequent to peripheral illumination. It was also established that preliminary adaptation of one eye to red light of low intensity heightens very markedly the peripheral sensitivity of both eyes (85–88; 117; 120; 143; 209; 216–218; 274, pp. 182–186; 286; 288; 289; 291; 292; 400; 402; 444; 472; 475).

Pronounced effects on the peripheral sensitivity of the visual organ have been demonstrated as a consequence of auditory stimulation (32, 78, 82, 137, 161, 178, 180, 183, 342–344, 345, 405, 446). For example, it has been shown that peripheral sensitivity, as a rule, declines on exposure to sounds of average or above average intensity. According to report, the noise of an airplane motor actually drops peripheral sensitivity to as low as 20 per cent of the level obtaining under conditions of quiet (32). However, a brief period of hyperventilation is held to restore peripheral sensitivity in spite of continued auditory action (94; 172, p. 118; 295; 296). Instances of heightened peripheral sensitivity have also been claimed (304, 319, 345), but later investigators have attributed these departures from the general rule to deviations from the subject's normal physiological state (137; 188; 274, pp. 186–187). However, exposure to ultrasonic frequencies (e.g., 32,800 cycles/sec.) has been reported to increase peripheral sensitivity (188).

As in the case of macular stimula-

tion, the same two-phase sequence of initial depression and subsequent elevation of peripheral sensitivity above normal—the latter phase a rather prolonged one—is to be discerned with auditory stimulation. This is true also when ultrasonic frequencies are employed (188; 274, p. 190).

Olfactory stimuli of various kinds have been observed to influence the level of peripheral sensitivity of the visual organ. The odor of bergamot oil and of pyridine in toluol, for example, are held to heighten peripheral sensitivity, although, where there were strong negative reactions to the latter, instances of lowered sensitivity are reported. Spirits of hartshorn are also said to increase peripheral sensitivity (48, pp. 281–286; 161; 183; 260; 274, p. 188; 345).

Similarly, gustatory stimuli of various kinds have been observed to influence the level of peripheral sensitivity of the visual organ. Weak sweet, salt, and acid solutions bring about an increase in peripheral sensitivity, whereas quinine brings on a decrease. The latter effect is, as above, attributed to the negative reaction of the subject to the bitterness of quinine (161, 186, 345).

Stimulation of the cold receptors is reported to produce a marked increase in peripheral sensitivity of the visual organ. For example, one minute after facial application of cool water, peripheral thresholds drop markedly. On the other hand, stimulation of the warmth receptors has a reverse effect, raising peripheral thresholds (66; 76; 172, p. 91; 183).

The effect of vestibular, intero-, and proprioceptive stimulation on peripheral sensitivity has been the subject of a number of studies. Interoceptive stimulation, arising from uteral expansion and distension of the bladder, is held to affect peripheral sensitivity adversely (172, pp. 109–

110; 185; 190; 191). Proprioceptive stimulation, evoked by muscular activity, is reported to induce varying effects: after light muscular activity, such as finger tapping at maximal rate or two minutes of gymnastic drill, an increase in peripheral sensitivity is noted; after heavy, fatiguing activity, however, a decrease ensues (34 40; 89; 91; 159; 172, pp. 91, 110–113; 185; 433). Postural stimuli exert varying effects on peripheral sensitivity of the visual organ: maximal sensitivity is elicited when the subject is comfortably seated; for the standing position it is considerably depressed (172, p. 118; 267). Vestibular stimulation—from rotation of a seated subject five or ten times—produces a significant decrease in peripheral sensitivity, with return to normality requiring periods ranging from 5 to 30 minutes (14, 96).

Dark Adaptation

Not only is peripheral sensitivity affected by various types of accessory stimulation, but so is the rate of dark adaptation. An accelerating influence on the process of dark adaptation has been reported from accessory stimulation of the cold, taste, and proprioceptive receptors; a similar effect upon accessory illumination of the eye with red light has also been reported (67; 157; 172, pp. 53–61; 288; 313; 316; 319; 401; 414).

Absolute Sensitivity of Central Vision

Considerable inquiry has been made into the effect of accessory stimulation on central visual sensitivity. From the research that has been done on this problem, it has been ascertained that auditory stimulation of moderate intensity heightens central sensitivity to white light for the dark-adapted eye, but that if monochromatic light is employed, the effect varies with the wave length

utilized. Thus, central sensitivity of the dark-adapted eye to blue-green colors is raised through auditory stimulation, whereas, to orange-red, it is decreased. Central sensitivity to extreme spectral red and violet, as well as to the yellow portion of the spectrum near 570 mu, remains unchanged however (172, p. 118; 225 227–234; 244; 250; 404).[8]

Kravkov, one of the foremost Soviet researchers in the field of sensory interaction, was able not only to demonstrate these effects over a considerable range of loudness, but also to show that the degree of effect varied directly with intensity of the accessory stimulus (244; 250. 274, pp. 270–273). He ascertained, further, that the observed changes in central sensitivity of the visual organ were a function of not only the intensity of the auditory stimulus but also of its duration. Thus, in an experiment to gauge the effect of sound on color sensitivity to green light, Kravkov found no effect from an auditory presentation of one-minute duration: however, he did note a pronounced effect on color sensitivity to green light following three minutes of auditory action.

The opposed character of color sensitivity to green and to red lights which is revealed when employing sound as accessory stimulus is disclosed also in experiments not involving auditory stimulation (263, pp. 35–39). For example, when the head is tilted back (postural accessory stimulation), it has been found that color sensitivity to green light (520

---

[8] Reverse accessory effects have also been noted on the central sensitivity of one eye when the other has undergone adaptation to red or green light. Adaptation of one eye to green light is reported to depress central sensitivity in the other upon accessory auditory action; adaptation to red light to elevate it (143).

mu. is always markedly decreased,
proceeding after 90 minutes to 25
per cent of its original value, whereas
color sensitivity to orange-red light
610 mu exhibits an occasional tend-
ency toward increase (267).

The direction of these effects is
the reverse of that elicited with sound
as the accessory stimulus. However,
chilling, thermal, and gustatory ac-
cessory stimuli have been reported to
induce directional effects similar to
those elicited by auditory accessory
stimuli. Here again, effects in the
yellow region and the spectral ex-
tremes are not to be observed (74; 76;
245; 251; 256, pp. 48–49; 424).

The opposed character of color
sensitivity to green and red lights is
demonstrated more fundamentally
by experiments of a quite different
type. Kravkov and Galochkina, for
example, were able to demonstrate
the opposing modifications of sensi-
tivity to green and red colors by pass-
ing a weak electric current through
the eye with strengths of current
ranging from 0.02 Ma. to 0.5 Ma.
and times of action up to 15 minutes.
Where anodal eye contact was em-
ployed, color sensitivity of the dark-
adapted eye was increased for green-
blue illumination, but decreased for
orange-red. On employing cathodal
eye contact, the effects were reversed.
No change was noted as regards the
extreme ends and yellow region of the
spectrum (274, pp. 281–282; 280;
283). The effect of auditory stimu-
lation on central sensitivity of the
eye was thereby shown to be similar
to that resulting from anelectrotonic
influence—a fact to which consider-
able significance is attached.

Differential Sensitivity to Brightness

Several experiments have been
performed which produced evidence
of an effect of accessory stimulation
on the differential sensitivity of the

eye to brightness 70 (pp. 216 220
98 99, 274 ;pp. 216 220 285 292
293). Illumination of one eye, it
seems, makes for a simultaneous de-
crease in differential sensitivity of
the other eye. Furthermore, the de-
crease is greater, the brighter the
white-lighted field on which the lat-
ter is focused. The same effect can
be obtained by keeping the bright-
ness of the test field constant and in-
creasing the brightness of accessory
illumination (85, 86, 88, 216, 218).

Similar effects are elicited with
sound as accessory stimulus. Thus,
the brighter the viewed field, the
greater the decrease in differential
sensitivity under the effect of simul-
taneous auditory stimulation (274,
p. 219); and very loud noise, such as
that produced by an airplane motor,
impairs markedly general brightness
discrimination (32).

Accessory stimulation of the pe-
ripheral and central areas of the
retina has also been shown to affect
subsequent differential sensitivities.
Thus, weak illumination separately
of the peripherally distributed rods
and cones acts to increase respec-
tively the differential sensitivities
of the centrally distributed cones and
rods, while adversely affecting re-
spectively those of their centrally dis-
tributed counterparts (274, pp. 220
221; 291–294).

Critical Flicker Frequency

Sensory interactional effects are
evidenced also in the results of stud-
ies on critical flicker frequency (c.f.f.)
under the influence of accessory stim-
ulation (21; 76; 219; 220; 246; 274,
pp. 357–370).[9] Kravkov, for example,
noted that illumination of one eye
with moderately bright light raises

[9] As was to be expected, accessorily in-
duced changes in photic sensitivities parallel
those noted in c.f.f., but with opposite sign
(274, p. 365).

c.f.f. in the other, but wards off the reduction of c.f.f. that is usually observed in central vision of the other eye in the course of dark adaptation (**274**, p. 363).

It was demonstrated, in addition, that the effects of auditory stimulation on c.f.f. depend on the monochromatic nature of the light used. Thus, during the time of auditory action c.f.f. for green light (520 mμ) is reduced, whereas for orange-red (630 mμ), c.f.f. is raised. Here central retinal illumination was of such brightness as to produce critical flicker frequencies of 12–20 flashes per second. Accessory stimulation was pitched at 800 cycles at 85 db. On employment of white light as the primary stimulus, accessory auditory action was likewise found to affect c.f.f.: for central vision it is heightened, for peripheral vision lowered (**220**; **246**; **256**, pp. 49–50; **274**, pp. 364–365).

Experimenters, employing the odors of bergamot oil and geraniol, ascertained the effect of olfactory stimulation on c.f.f. (**251**; **274**, pp. 273–275). With brightness of flicker acting upon the central retina so as to produce a c.f.f. of 12–18 flashes per second for an eye at a given level of dark adaptation, the reduction of c.f.f. for green-blue illumination and its increase for orange-red under olfactory stimulation were established. No effects were noted for yellow (around 570 mμ) or for extreme spectral red (690 mμ) and violet (425 mμ). Furthermore, upon cessation of accessory action, it was found that c.f.f. returned not only to previous levels, but having done so, frequently proceeded to exhibit relative frequencies the reverse of those previously obtained under olfactory influence. A similar picture of change in c.f.f. has also been disclosed upon accessory application of gustatory and thermal stimuli (**76**).

Reverse effects have been obtained by varying the intensity of light employed. Thus, if brighter light should be employed to bring the corresponding c.f.f. to 26–30 flashes per second, then under the action of the same accessory stimulation c.f.f. may change in the opposite direction to that obtaining for brightness marked by a c.f.f. of 12–18 flashes per second. In other words, if c.f.f. for weak light is reduced by accessory stimulation, then much brighter light may under the same conditions increase it (**274**, pp. 367–368).

Visual Acuity and Irradiation

A number of studies have been devoted to research on the effect of accessory stimulation on visual acuity (**119**; **209**; **210**; **214**; **215**; **233**, p. 472; **256**, p. 50; **265**, pp. 26–29; **276–278**). These show, for example, that where closely set, black forms are to be distinguished from each other against a white background, visual acuity is sharpened under accessory stimulation; but where the shades are interchanged, it is diminished. Of interest also is the fact that, upon monaural stimulation, especially marked effects are to be noted on the visual acuity of the contralateral eye (**274**, p. 187).

In irradiational terms one can speak of the increase of irradiational effect under accessory stimulation. Visual acuity in the above case may then be thought of as enhanced because, through increased irradiational effect under accessory stimulation, the white interspaces between the black forms crowded on a white background are rendered larger and the black forms are consequently better brought out. Where the shades are interchanged, the reduction of the black interspaces under accessory stimulation would explain the resulting diminution of visual acuity (**214**, **215**, **276–278**).

Whether the accessory stimulation be through light, sound, or odor, the increase of irradiational effect for white is demonstrable in every case. For example, increase of irradiational effect is induced by the odor of bergamot oil (214, 251), and under accessory auditory stimulation irradiational effect for orange-red increases (207, p. 34)—a finding which might not have been anticipated in view of the reputed decline of color sensitivity to orange-red under auditory stimulation. This fact is considered to be of importance in surmising a probable mechanism behind the action of accessory stimuli on visual afferentiation, and we will return to it later.

Electrical Sensitivity of the Eye

Electrical sensitivity of the eye[10] has been shown to be modifiable by the action of accessory stimuli (30, pp. 162-164; 70, p. 188; 94; 139; 410; 412; 415; 462; 463; 479). For example, illumination of one eye always raises the electrical sensitivity of the other. Auditory stimulation likewise increases electrical sensitivity, provided the sounds employed are not extremely loud, in which case electrical sensitivity is decreased (20, 30, 130, 132). Gustatory stimulation produces varying effects on the electrical sensitivity of the eye. Where sugar furnishes the accessory stimulus, electrical sensitivity is increased; where the stimulus is salt or citric acid, it is reduced. Under the olfactory action of bergamot oil increase in threshold of electrical sensitivity is reported to result (71, 72).

Color Zones

The boundaries of color zones have been reported to shift under accessory stimulation. Thus with intense stimulation of the retina with bright light expansion of the red and yellow color zones and contraction of the green and blue zones result (105). With sound as accessory stimulus the zonal boundaries for green and blue are enlarged, those for orange-red drawn in, while those for extreme red remain as they are (144-146)—effects which are reminiscent of the changes, previously described, in color sensitivity brought on through accessory auditory stimulation.

Olfactory stimuli also have effects on the zonal boundaries (48, p. 281). If the odors of rosemary and indol are employed, expansion of the green region results with the former and its contraction with the latter. As regards the red zone, rosemary is found to contract, but indol to expand it—effects the reverse of those above. However, alteration of red zonal boundaries is not always observed (274, p. 275; 398; 399, pp. 389-396).

Contrast

Changes in visual contrast effects under the influence of auditory stimulation have also been the subject of investigation (446). The effects vary with the degree of contrast obtaining prior to accessory action. Thus, if a gray circular area on a white background is presented with a tone of 800 cycles at 75 db as accessory stimulus, contrast is heightened if the initial difference of brightness between the gray circle and white background is in itself great. On the other hand, contrast is diminished if the difference is initially small.

Afterimages

Accessory auditory stimulation has been shown to affect the course of extinction of visual afterimages (363, 364, 469-472). In a number of

[10] Electrical sensitivity of the eye is inversely indexed by the minimum strength of the applied current necessary to induce visual sensation (phosphene) on closing and breaking the circuit. Thus, the weaker the current inducing phosphene, the higher the electrical sensitivity of the eye.

people, for example, a strong auditory stimulus heightens the brightness of the afterimage, though hastening the termination of its course. Here the accessory stimulus precedes the inception of retinal illumination and is continued until extinction of the resulting afterimage. Sounds that are powerful enough to evoke fairly intense auditory afterimages tend especially to heighten the brightness of the visual afterimage. However, should an extremely loud tone be presented after the cessation of retinal illumination, the afterimage undergoes temporarily complete disappearance.

Auditory Sensitivity

While most research has employed light as the primary stimulus, a number of studies have been undertaken with sound performing the role. Here again effects of accessory stimulation have been demonstrated. Illumination of the eyes with white light, for example, is claimed to increase auditory sensitivity; absence of visual stimulation, on the contrary, decreases it. However, opposed effects are obtained by using, as accessory stimuli, different monochromatic lights. Thus, illumination of a white room with green light increases auditory sensitivity; but with red light it is decreased (**182; 202; 312,** p. 155; **351; 429**).

Auditory thresholds are lowered also when the odors derived from geraniol and benzol supply the accessory stimulation (**48,** p. 281; **134; 267; 379; 380**). Postural stimuli likewise are reported to affect auditory sensitivity. Thus, on tilting the head backward for any length of time auditory sensitivity is found to suffer. It is also reported that higher auditory sensitivity is obtained in the seated position than in either the standing or horizontal position (**172**, p. 118).

Gustatory stimulation has been reported to improve auditory reception of low tones, but to worsen somewhat that of high tones. Established also have been the facilitating effects on hearing of accessory stimulation of the cold and proprioceptive receptors (**172**, p. 82; **366**).

Miscellany

Comparatively little work has been done where the primary stimulus has not been visual or auditory (**327**, p. 763). But what work has been done shows the presence of accessory effects. Thus, it is averred, illumination of the eyes with white light heightens gustatory and olfactory sensitivity (**48,** pp. 253–254), whereas auditory stimulation reduces the electrical sensitivity of the gustatory apparatus (**194; 267,** p. 50). Electrical sensitivity of the gustatory apparatus, as well as its sensitivity to salt solution, is reported also to decrease in the course of dark adaptation of the eye. Mental activity, however, excludes the latter effect (**48,** pp. 140–141; **71; 72; 457**).

Research on the accessory effect of the odors of geraniol and thymol on the vestibular apparatus discloses reverse effects. The odor of geraniol heightens vestibular excitability, prolongs rotary postnystagmus, lengthens the illusion of counterrotation, and shortens chronaxie, while that of thymol has the opposite effect (**48,** pp. 281–282; **150; 151**).

SOME CONDITIONS AFFECTING SENSORY INTERACTION

Strength of Accessory Stimulus

Increase in strength of the accessory stimulus leads so frequently to effects that are the reverse of those induced by weaker intensities that

one can almost speak of a "rule of inversion" (67; 111; 161; 172, pp. 80, 88; 178; 184; 226; 263, pp. 39-41; 274, pp. 180-181, 190; 344; 448-451; 474). For example, sounds of weak intensity heighten the electrical sensitivity of the eye, while those of increasing intensity are accompanied by a gradual decline. However, Kravkov doubts the generality of any "rule of inversion" and points to one case where the rule itself is controverted. Varying the intensities of sounds employed as accessory stimulation, he found that an increase from 25 db to 95 db was paralleled by a steady increase of sensitivity to green light (530 mμ) and a steady decrease to orange light (590 mμ)—findings which deny a general "rule of inversion" (267, p. 93; 274, p. 180).

Excitatory State of Primary Sense Organ

The effect of accessory stimulation depends not only on the intensity of the accessory stimulus, but also on the initial degree of excitation of the sense organ undergoing primary stimulation. Thus, in order to increase the loudness of a tone on accessory visual stimulation, the initial loudness must be sufficiently great. Otherwise, either no effect will be demonstrable or a loss in loudness will ensue. The example cited under *Contrast* in the previous section is another instance of the effect of initial degree of primary excitation.

Duration of Accessory Stimulation

Intensity of accessory stimulation is not the only factor of importance; duration is also important. Kravkov (130-132, 184, 344, 348) has shown that, as a rule, accessory effect becomes increasingly prominent during the first moments of accessory action. But subsequently the effect very of-

attained (274, p. 186). For example, auditory stimulation was underwent simultaneously 31 to 61 minutes of dark adaptation. Critical flicker frequency was found during this period to rise sharply, but to diminish after a quarter hour from (220).

Termination of Accessory Stimulation.

In describing the influence of accessory stimulation on a given sense organ, it is, of course, necessary to stipulate whether the state of the latter is tested before or after cessation of accessory stimulation since, after withdrawal of the accessory stimulus, its effect does not just wear off with an eventual return to normalcy. There frequently arises instead a shift in a direction opposite to that of the initial effect. Lowered sensitivity may be followed by a supernormal phase, heightened sensitivity by a subnormal (157, 274, p. 190).

For example, Kravkov observed that sound lowers c.f.f. for peripheral vision. After its termination, however, he noted that under the same experimental conditions c.f.f. is raised, sometimes markedly (220; 274, p. 364). As a further illustration, another experiment may be cited (405). Here, on testing the peripheral sensitivity of the eye after cessation of auditory stimulation, it was observed that a supernormal phase emerged subsequent to the decrease generally registered during the action of the accessory stimulus.

Affectivity of Stimulus

It appears that stimulus affectivity is a factor to be reckoned with in sensory interaction. For example, it has been claimed that, where gusta-

tory and olfactory stimuli of unpleasant character are applied as accessory agents, peripheral sensitivity of the eye declines (172, p. 88, 345). It is also reported that, after presentation of harmonious tones, color sensitivity of the dark-adapted eye to red and yellow is raised, to green-blue lowered, whereas disagreeable sounds reverse the effects (259, 426, 428).

Physiological State

The specific physiological state of the sense organs, as well as the more general physiological state of the individual, plays a role in sensory interaction (274, p. 169; 399, pp. 412–421; 455).[11] A number of experiments may be mentioned by way of example. In a previously cited experiment, it was shown that presence of a subject in a green-illuminated white room heightens auditory sensitivity, whereas presence in a red-illuminated white room brings about a decline. However, on changing the physiological state of the subject through moderate dosage of veronal, the experimental effects are precisely reversed. In the hyposthenic state induced by veronal, green illumination lowers, red illumination heightens auditory sensitivity (189, 429).

Similarly, it has been demonstrated that with intense auditory stimulation the electrical sensitivity of the light-adapted eye undergoes decrease, and that of the dark-adapted eye, increase (20; 139; 274, p. 188; 303). Electrotonic stimulation of the eye with anodal contact thereon (anelectrotonus) induces similar effects; with cathodal contact (catelectrotonus), however, reverse effects show up (274, p. 190; 411).

Instillation of adrenalin into the

[11] In various hypodynamic states, for example, the more usual effects of accessory stimulation tend to sharper expression (172, pp. 97–117).

eye has likewise been shown to bring on similar modifications in the electrical sensitivity of the eye: for the dark-adapted eye, electrical sensitivity declines; for the light adapted it increases (106, 303). Similarly various accessory effects on electrical and color sensitivity can be made to reverse sign depending on whether the tested eye is light- or dark-adapted or whether it is adapted to green or red light (235, 240, 249, 268, 298, 415).

Miscellany

Summation of accessory effects. If each of two sense modalities, on accessory stimulation, induces an increase in peripheral sensitivity of the eye, their joint action induces an effect beyond that induced singly. If, however, the individual effects are in opposite directions, their joint action induces a general effect on peripheral sensitivity equal approximately to their difference (172, pp. 80–81; 181).

Repetition of accessory stimulation. Constant repetition over equal intervals of time of the same accessory stimulus has been found frequently to yield a diminishing effect which tends to zero (172, p. 81).

Cumulation of accessory effects. The effects of accessory stimulation linger in some instances so that on subsequent days the change in level of primary threshold is found to undergo successive increase (172, p. 81).

Diurnal variation. The respective sensitivities of the rod and cone apparatus exhibit an inverse reciprocity of accessory influence that is diurnally dependent. Thus, it has been shown that at midday, when central sensitivity is at its maximum, peripheral sensitivity is minimal. As a normal relation, degree of sensitivity of the one apparatus is inversely tied to

degree of sensitivity of the other (431, 436–438).

SOME THEORY ON SENSORY INTERACTION

In the attempt to comprehend the nature of the modifications of sensory response induced through accessory stimulus action, several considerations have been put to theoretical work. The result to date appears more a collection of eclectic plausibilities than an integrated set of relevant explanations. The following discussion will report what these considerations are.

Contiguity

Over a considerable range of auditory and visual stimuli certain visual and auditory responses undergo intensification through respective accessory reinforcement via intermodal action. To explain a number of them, such as the increase in visual acuity of one eye upon monaural stimulation of the contralateral ear, recourse has been had to the neurological map. In the region of the *corpora quadrigemina*, for example, the visual and auditory nerve fibers are in close proximity and myelin-free. Since nerve action in one unmyelinated fiber is held likely to influence that of another contiguous to it, the feasibility of excitatory irradiation through contiguity of the conducting fibers is considered to be demonstrated. The fact and feasibility of excitatory irradiation in the region of the *corpora quadrigemina* and also of the *corpora geniculata* seem to be justified, in the main, by reference to foreign work (6; 121; 152; 156; 265, p. 29; 274, pp. 187–188; 304; 319).

Similar considerations are thought to make olfactory influence on vision understandable, for in the region of the *stratum zonale* and *nucleus anterior thalami* the visual and olfactory nerve fibers lie in close proximity. Thus the olfactory afferent system may likewise have its effect on visual phenomena through excitatory irradiation.

The fact that there is anatomical proximity along the auditory and visual nerve pathways is also thought to make explicable many cases of synesthesia. On this basis the latter can be viewed as instances of sensory interaction (267, pp. 59–63).

Neural Excitation

The question arises as to whether it is better to view many instances of accessory effect as mediable through change in neural excitation rather than through change in neural excitability, that is to say, through threshold modification, although either alternative could conceivably account for a number of the intensifying effects produced by accessory stimulation. Kravkov and others, as a matter of fact, designed a number of experiments to throw light on just this point (265, pp. 26–29; 274, pp. 222–230).

It had been established earlier that the sensitivity of central vision to orange-red light declines with simultaneous auditory stimulation; that is to say, a rise of threshold or decrease of excitability takes place. Kravkov reasoned that, if on accessory auditory stimulation positive irradiation could, nevertheless, be demonstrated to intensify for an object of orange-red hue (if its apparent size was increased), then change not in neural excitability, but in neural excitation must be decisive for achieving the observed effects. If positive irradiation decreased, then change in neural excitability could be taken to operate as the major factor.

Experiment realized the former contingency—the effect of positive

irradiation was increased, a result leading to the conclusion that accessory auditory stimulation may add to visual excitation of certain kinds (257). Thus, on this basis, the data, relevant to both positive irradiational effects and visual acuity on accessory auditory stimulation, are held explicable.

Further support for the emergence of an accessory effect through change in neural excitation is derived from the fact that with accessory auditory stimulation increase of differential threshold is not accompanied by a seemingly necessary decrease of irradiational effect. It is held that this failure of prediction is also accounted for on the assumption of an accessorily induced upward modification of neural excitation which renders $\Delta C/C$ greater in magnitude than $\Delta I/I$, of which it is the "central physiological correlate, . . . C and ΔC being measures of excitation in the nervous centers" (274, pp. 222–225).

Leveling and Accentuation

Under the "rule of leveling and accentuation" a refinement, implicit in the foregoing, was explicitly introduced to allow a wider encompassing of disparate data (257; 263, pp. 31–35; 265, pp. 26–29; 274, pp. 225–230). This refinement consists in taking note of the initial contrast between the viewed object and its background. The following discussion points to this necessity.

It was shown that, with simultaneous auditory stimulation, visual irradiation of a white object against a black background intensifies, while that of a gray object against the same background diminishes (215). Here is an example of the kind of oppositely directed effects for which a unified explanation is needed. The rule of leveling and accentuation that has

been proposed aims to provide one where initial contrast is large, accentuation of contrast occurs; where it is small, leveling or reduction of contrast results. The reasoning behind the rule is fairly direct and may be indicated by way of example.

Supplementary excitation of the visual apparatus on accessory auditory stimulation is not distributed uniformly over the visual field but is unevenly apportioned. Those regions of the visual apparatus which already are to a considerable degree excited, "attract to themselves" the major portion of the additional excitation; the less strongly excited regions share more modestly. In other words, the magnitude of supplemental excitation, accessorily induced, is a function of the excitation directly occasioned by the primary stimulus. Thus, on accessory visual stimulation loud tones should sound louder; weak tones should be little, if at all, affected. In this way the rule of accentuation is held explicable.

But why a rule of accentuation and leveling? Again an example will elucidate. It is known that differential thresholds for brightness discrimination increase with brightness. If there is little initial brightness contrast between object and background, accessory increase of excitation is about equally shared, so that, where the ratio of differential brightness may have been I_2/I_1, it is centrally now $I_2+\Delta E/I_1+\Delta E$; ΔE is taken as the "brightness equivalent of the supplementary excitation" and $I_2 > I_1$. This leaves the resulting difference in excitation about the same. The magnitude of the latter ratio, now less than the former, indicates, however, an impairment in sensing the brightness difference that obtains. Thus, brightness difference between object and background, initially contrasting little, comes to be even harder to ob-

serve with the result that a leveling tendency in brightness difference establishes itself (**274**, p. 225).

On the the other hand, should the initial contrast be very large, accessory increase of excitation accrues, in the main, to the brighter source as previously explained, so that the effect of contrast comes in this instance to be seen as exaggerated (**353, 446**).

Similar reasoning would account for the oppositely directed shifts in critical flicker frequency that are to be observed when accessory auditory stimulation is employed to reinforce flashes of low and high brightness respectively. Critical flicker frequency is increased when the brightness of flash is initially high; it is decreased when initially low. Here the flash is thought of as an object presented against a less bright background represented by the fading afterimage between flash phases (**274**, pp. 366–370).

Intracentral Mediation

An accumulation of experimental data (**274**), as well as the anatomical richness of intracentral connections, made natural the consideration of intracentral mediation as a factor in sensory interaction. Certain specific research seemed also to suggest the intracentral mediational character of accessory action. For example, in the course of registering the bioelectrical currents in the auditory region of the cerebral cortex of rabbits, it was discovered that these currents are modified in an interesting fashion, should the eye of the animal be exposed to regular flicker: the rhythm of electrical potential in the auditory region comes, after an interval, to correspond to that of the flicker (**332**). In addition, the facts of protopathic and epicritic sensitivities were taken also to demonstrate

the presence of central inhibitory action of some afferential systems on others. Thus, the tactile and proprioceptive systems, taken as phylogenetically of more recent origin, are seen as antagonistic in action to the systems, very much those sensations of gross temperature and other pain which appear on disturbance of the inhibitory action of the former 265, p. 29; **327**, p. 611; **371**; **372**, pp. 39–53; **375**).

The widespread influence of the Pavlovian propositions of cortical action (since 1950 offered dogma) has also had its effect in increasing the plausibility of intracentral mediation of accessory action 274, pp. 178–180, **338–340; 374**, pp. 136–296). Thus, on the assumption of such mediation two possibilities present themselves: either excitation of a cortical center leads to excitation of another, medially distinct, or it leads to depression of the latter. Accessory effects may, therefore, be reinforcing or antagonistic.

It has previously been pointed out that central illumination raises peripheral threshold, while peripheral stimulation decreases central sensitivity. The relationship is mutually antagonistic, provided central vision is not that of the achromat, in whom an antagonistic influence is not exhibited due to cone deficiencies. Since central illumination of one eye, for example, affects adversely peripheral sensitivity in the other, it has been concluded that such effects, as indicated above, must, therefore, be intracentrally mediated (**391**).

Intracentral mediation of accessory action functions also to account for many of the reverse effects elicited upon increase in strength of the accessory stimulus ("rule of inversion"). The assumption is made that an initially weak accessory stimulus sets up two central proc-

esses: one positively inductive; the other negatively inductive. Upon increase in strength of the accessory stimulus, increases in the magnitude of the two oppositely signed inductive effects are given by two differently accelerated growth curves. Reversal of modal accessory effect is associated, therefore, with the change of sign resulting from algebraic summation of the two curves, once a certain magnitude of the accessory stimulus has been exceeded (**274**, p. 181).

Ionic Balance

It has been pointed out that under anelectrotonic influence central sensitivity of the dark-adapted eye to green-blue is increased, to orange-red decreased, whereas under catelectrotonic influence the reverse is observed. These changes parallel remarkably those resulting from the application of appropriate accessory stimuli and, hence, have been the subject of special study.

Attempts to explain the results of electrotonic influence have generally been based on considerations of ionic concentration and balance (**274**, pp. 281–286; **280**, pp. 138–143; **415**). In view of the known role of ionic calcium and potassium in the excitability of tissues, investigation turned first to them and later to ionic sodium and magnesium.

Since in a solution potassium ions are univalent and those of calcium bivalent, the former migrate to the cathode more rapidly than the latter upon application of direct current to a circuit. The value of the ionic index [K]/[Ca], which measures the relative concentration of ionic potassium and calcium, increases, therefore, at the cathode; it decreases, of course, at the anode due to increasing concentration of ionic calcium there.

If ionic factors are involved in the reported modifications of central sensitivity, it was reasoned that changes in the relative concentration of ions should be accompanied by changes in the various central sensitivities. Through iontophoresis[18] it was possible to introduce ionic calcium and potassium into the eyeball in order to test the surmise. Predictions were fulfilled (**274**, p. 284; **283**): ionic calcium was found to increase sensitivity to green but to decrease it to orange-red, whereas ionic potassium was found to increase sensitivity to orange-red but to decrease it to green. Termination of the iontophoretic process reversed the respective effects (**262**, **281**, **282**, **284**).[19]

Further experimentation showed that ionic sodium was similar in action to ionic potassium, while ionic magnesium was similar in action to ionic calcium (**260**, **376**, **377**). Hence, decrease in value of the ionic index $[K]+[Na]/[Ca]+[Mg]$ is associated with increase in sensitivity of the green- and blue-sensing apparatus of the eye, while increase in value of the ionic index is tied to increase in sensitivity of the red-sensing apparatus.

Since ionic sodium and potassium are known to favor the concentration of acetylcholine in tissues, whereas ionic calcium and magnesium are known to lead to its reduction, it is considered very likely that various concentrations of acetylcholine exhibit respective optimal effects for the various color-sensing apparatuses of the eye (**274**, p. 284). In the last

[18] The introduction of ions into the body by means of electric current.

[19] Through iontophoresis it was found possible to neutralize certain accessory effects on central sensitivity. Thus, iontophoresis of sodium, which reduces central sensitivity to green (520 mμ), neutralizes the accessory effect of auditory stimulation which, under ordinary conditions, heightens this sensitivity (**275**, p. 146).

analysis, then, central sensitivities are seen as dependent on ionically influenced acetylcholine concentration (275, pp. 143–145).[14]

The Autonomic Nervous System

A variety of accessory stimuli such as tones, noise, the taste of sweetness, the odor of bergamot oil, geraniol, and camphor has been reported to produce similar differential effects on retinal sensitivity to green-blue and orange-red lights: heightening sensitivity to the former, decreasing sensitivity to the latter (228, 234, 251). This circumstance led researchers to look for a common factor behind this discerned uniformity of differential action. They noted that application of the above stimuli was accompanied by a quickening of the pulse, and it was conjectured that the common factor might well be the autonomic nervous system, in view of the fact also that the odor of thymol, which reverses the differential effects cited above, was found to reduce pulse rate (8; 9; 160; 172, pp. 81, 88; 242; 252; 253; 256, p. 49; 265, p. 31; 399, pp. 412–419).

This conjecture among others led to experimentation which disclosed that the color sensitivity of the eye was changed in the reciprocal pattern mentioned above on direct instillation of adrenalin into the eye. Further research along these lines demon-

strated that sympathomimetic substances such as cordiamin, phenamine, sympathol, and ephedrine heighten sensitivity to green light for the dark-adapted eye, but lower it to red light. On the other hand, parasympathomimetic substances such as carbocholin, veronal, the berries of *Schizandra Chinensis*, and pilocarpine usually bring on reverse effects: sensitivity to red light is heightened for the dark-adapted eye, while that to green light is lowered. Ephedrine is reported to heighten also gustatory, olfactory, and auditory sensitivities (48, pp. 255–258; 112; 114; 115; 147; 148; 153; 252; 253; 274, pp. 276–278; 279; 303; 399, pp. 401–407).

Confirmation of the autonomic mediation of a number of accessory effects has been sought also in data from experiments involving electrical application to the eyeball. For example, it will be remembered that if direct current of weak intensity is passed through the dark-adapted eye with anodal contact, sensitivity to green-blue light is heightened with corresponding diminution to orange-red light—the characteristic antithesis. The effects on sensitivities are reversed with cathodal contact (268; 274, p. 190; 280; 283; 411).

Now, it was pointed out that the relative concentration of calcium ions is increased in the vicinity of the anode during electrical action. Since increase in relative concentration of calcium ions, in a very great number of instances, is understood to induce effects similar to those of sympathetic excitation, the mechanism of the above accessory effects is held to be probably autonomic (262; 265, p. 33; 274, pp. 285–286; 282; 284).

Therefore, auditory stimulation, the odor of bergamot oil and geraniol, and the taste of sweetness are said to achieve their accessory effects through influence on the sympathetic

---

[14] Ionic calcium and magnesium are known to activize cholinesterase which through hydrolysis reduces acetylcholine content; ionic sodium and potassium, on the other hand, are known to suppress cholinesterase, thereby increasing acetylcholine content.

Interference with the acetylcholine chain of reactions, as through homatropinization of the eye, stabilizes central sensitivities in spite of iontophoresis. Hence, strictly speaking, not acetylcholine, but processes involving it in an essential way are to be associated with modifications of central sensitivities (354, 377).

division of the autonomic nervous system with differential effects on the color-sensing apparatus of the eye; adverse effects are induced where parasympathetic excitation would instead ordinarily facilitate.

Corroboration of autonomic action has been sought in a variety of experiments. For example, backward tilting of the head for a period of time is claimed to bring about conditions favoring parasympathetic excitation. One should then expect a reduction of sensitivity to green light—an expectation which was confirmed (265, p. 32; 423).

It should be noted that the differential action of the two divisions of the autonomic nervous system on the green- and red-sensing apparatus is also consonant with the observed reciprocal action of green and red light on a number of physiological functions. For example, the internal ocular pressure of one eye is increased or decreased according as the other eye is illuminated with red or green light respectively (476, 477). A similar reciprocity of action has also been observed with hyperventilation which decreases sensitivity to green light while increasing it to red (296).

Certain changes in peripheral vision are likewise held ascribable to modifications of the autonomic state (8, 9). Thus, vestibular stimulation, which is said to favor parasympathetic excitation, was noted to worsen peripheral sensitivity (14). Backward tilting of the head is also reported to decrease peripheral sensitivity, which fact accords with the increase of parasympathetic excitation held attributable to the tilted position of the head. Changes in peripheral sensitivity have also been noted with alteration of relative ionic concentrations, electrically induced—a fact which again is taken to relate modification in peripheral

sensitivity to autonomic action (283).

Further confirmation of the autonomic mediation of accessory action was sought in experiments involving the cerebellum. For example, it was shown that subjection of the cerebellum to electromagnetic waves of ultra high frequency definitely altered the level of sensitivity in peripheral vision (334). Since according to Orbeli, the cerebellum must be regarded as one of the major regulators of sympathetic nervous activity, the results of experiments such as the above were taken as tending to demonstrate the partial dependence of the visual apparatus on the autonomic nervous system (372, 383).[15]

While a wide variety of data is accounted for on the basis of autonomic mediation of accessory action, Kravkov points out that the initial drop in peripheral visual sensitivity on accessory auditory stimulation is an experimental datum which does not harmonize with all of the foregoing. Auditory stimuli appear to be sympathotropic as to effects: pulse rate is stepped up, the sensitivity of the green-sensing apparatus of the eye is heightened, the retinal blood vessels are constricted, etc. But peripheral sensitivity appears to be sympathotropically affected inasmuch as it increases under anelectrotonic conditions and decreases under vagotropic stimuli such as are provided by the backward tilting of the head. One should, therefore, expect a positive effect on peripheral sensi-

[15] Orbeli's theory of the "adaptive-trophic" regulatory action of the sympathetic nervous system on all tissues of the body, including the central nervous system and the striped musculature, has had a considerable influence on research in the field of sensory interaction (55; 69; 95; 123; 160; 162, p. 77; 173; 192; 193; 256, p. 46; 334; 336; 337; 368-374; 456; 468).

tivity through accessory auditory
stimulation. The reverse, however,
takes place. To account for this,
Kravkov hypothesizes the existence
of intracentral connections whose in-
fluence in an *ad hoc* fashion domi-
nates antagonistically over those
which is proceeding through the
medium of the autonomic system
(267, p. 81).

The Green Receptors

The fact that the application of
either sympatho- or vagotropic ac-
cessory stimuli leaves unaffected
sensitivity both to extreme spectral
red and violet and to colors in the
region of 570 mμ has led to the assign-
ment of an especial role to the green-
sensing apparatus of the eye. It is
hypothesized that only if the green
receptors are stimulated is accessory
heightening or lowering of color sensi-
tivity evocable (254; 256, p. 49; 266;
274, pp. 327–331; 352).

Throughout the spectral range of
excitability of the green receptors
(from 430 mμ to 680 mμ approxi-
mately) excitation of the green-
sensing apparatus appears every-
where accompanied by accessory ef-
fects except in the neighborhood of
570 mμ. Reference to the three exci-
tation curves of the three basic color
receptors (b, g, r) shows that the ordi-
nates Y_g and Y_r are about equal in
the neighborhood of 570 mμ, Y being
taken as equal to degree of excitation.
Since excitation of the green receptors
is held to induce a positive accessory
effect, the failure of accessory effect
in the neighborhood of 570 mμ is
ascribed to an initiated negative ac-
cessory effect of equal magnitude, in-
duced by a simultaneous excitation
of the red receptors.

Since magnitude of unsummated
accessory effect is regarded as a
direct function of degree of excitation
of the respective color-sensing ap-

paratus is realized where $Y_g > Y_r$, the
algebraic sum of the summated posi-
tive and negative accessory effects
yields a residual accessory effect
which is positive; where $Y_g < Y_r$, the
residual accessory effect is negative;
where $Y_g = Y_r$, accessory effect is
lacking (570 mμ, 680 mμ approximate-
ly).

Excitation of the blue receptors is
thought to enhance accessory effect
on the condition that mμ > 430 ap-
proximately. Thus, in the neigh-
borhood of 520 mμ where $Y_b = Y_r$
approximately, the inhibitory action
of the red receptors is counterbalanced by
the sensibilizing action of the blue
receptors. This produces, accordingly, a positive accessory effect derived
solely from stimulation of the green
receptors, one which, as it turns out,
is maximal.

On making the inhibitory and sen-
sibilizing action of the red and blue
receptors respectively explicitly de-
pendent upon excitation of the green-
sensing apparatus, the failure to elicit
accessory effects at the spectral ex-
tremes (mμ < 430; mμ > 680 approxi-
mately) is made understandable.

To test the hypothesis that ac-
cessory effects on color sensitivity in
trichromats are not observed without
excitation of the green-sensing ap-
paratus, an experiment was per-
formed to see whether through cen-
tral mediation accessory effects at
the spectral extremes could not be
elicited in one eye upon illumination
of the other with green light. The
results were affirmative, thereby con-
firming the special role assigned to
the green receptors (253, 254, 352).

Modification of Primary Conditions

Accessory stimulation is held to
produce its effect in many instances
in ways not involving considerations
of excitation or excitability of the
primary sensory system. These refer

to accessorily induced modifications of the physical conditions under which primary stimulation is applied. For example, the pupillary reflex may respond to an auditory stimulus. Here the quick contraction of the pupil with its slow subsequent relaxation would be expected to modify some aspects of the visual function. Since the amount of light entering the eye is affected by pupillary width, upon nonemployment of an artificial pupil apparent changes in visual threshold which follow upon auditory stimulation may be ascribable not to changes in excitation or excitability of the visual afferent system, but rather to changes, induced by accessory stimulation, in the physical conditions. Since the pupillary reflex is not without effect on accommodation, visual acuity also may undergo modification owing to the accessorily induced changes in the physical conditions accompanying primary stimulus action.

Similarly, pain, vestibular stimulation, etc. modify the pupillary reflex and thereby may give rise to simulated accessory effects. As another instance, tactile stimulation around the ear appears to affect auditory threshold; but again, the result of such stimulation might more aptly be ascribed to reflex changes in the small muscles controlling the tension of the bony system of the middle ear than to changes in the auditory afferent system (267, pp. 87–88; 274, pp. 86–88).

Conditioning

The conditionability of sensory responses adds immensely to the range of sensory interaction and is considered a factor to be reckoned with though the accessory effect is not native but acquired. Work in the Soviet Union on conditioned "sensory reflexes" dates back to 1936 when Kravkov, Kekcheev, Bogoslovkiĭ, and Dolin independently demonstrated the conditionability of changes in visual responses. Thus, it was early shown possible to condition changes in electrical sensitivity of the eye, critical flicker frequency, peripheral sensitivity, etc. to such indifferent stimuli as ticking, a tone, and elapsed time of eye exposure to the dark (16–18; 22–25; 50–52; 57; 79–84; 124; 169–171; 183; 202; 233, p. 474; 256, p. 53; 259; 263, pp. 41–43; 265, pp. 36–39; 267, p. 35; 274, p. 192; 309; 339; 381; 396; 417; 418).

To illustrate some of this work: Pshonik managed to induce changes in thermal sensation in response to accessory auditory stimulation by making of the latter a conditioned stimulus (57, 381). Dobriakova conditioned increase of gustatory sensitivity to the ticking of a metronome after having associated the latter with light—a stimulus which ordinarily induces a positive accessory effect on gustatory sensitivity. In addition, it was found that symbols, denoting stimuli which heighten gustatory sensitivity, make manifest like effects (71, 72, 171). A general heightening of various visual and auditory sensitivities has been observed to ensue through mere presence of a subject in the experimental room. Frequent prior presence in the room while in a state of intense attention is here said to condition modification of visual and auditory sensitivities to the sight of the experimental room—the state of attention being thought of as in the role of "unconditioned stimulus" (79; 80; 267, pp. 85, 111; 272; 274, pp. 192–194).[16]

[16] "Concentrated attention" has been assigned an important role in the sensibilization of the receptors. "Attentive practice" and expectation also lead to sensibilization, the former being held capable of doing so on a permanent basis (48, p. 209, 149, 172, p. 92; 274, pp. 230–231; 396; 408; 416; 460).

CONCLUDING REMARKS

Soviet research on sensory interaction has amassed over the past 24 years a fund of data that, in spite of a deserved skepticism, cannot but strike one for its persistent consistency and "logical," though novel, character. This research appears to demonstrate that all modalities undergo various modifications of sensory response on appropriate application of an accessory stimulus and that, where the primary stimulus is visual, the resulting modifications conform to striking patterns.

Basic to one of these patterns is the fundamental role assigned to the green-sensing apparatus whose excitation is a requisite for the mediability of accessory effects through simultaneous excitation of the non-green receptors. Of the latter there are those which upon excitation are reinforcing for accessory effects, while others are depressive. Thus, where sensitivity to blue-green, for example, is heightened, that to orange-red is lowered; and vice versa.

The antagonistic relationship between color receptors is traced to inversely related ionic concentrations induced through autonomic nervous action brought into accessory play.

Thus, where ionic calcium predominates, sensitivity to green is heightened, to orange-red lowered; where ionic potassium predominates, reverse sensitivities obtain.

Other patterns of accessory effect and of reciprocal accessory action have been intensively investigated with characteristic results. However critical one may be of the Soviet studies, the cumulative results obtained with their apparent inner consistencies suggest, therefore, the worthwhileness of a re-examination of the whole problem of sensory interaction. Recent restrictions on types of psychological research in the U.S.S.R. in favor of "Pavlovian" methodologies and conceptions suggest that non-Soviet researchers are the ones to undertake the job (60; 201; 272; 338–341; 434, pp. 69–70; 473; 498–503; 505).[17]

---

[17] Signs point to a possible relaxation of dogmatic prescription of allowable methodologies and conceptions. Both Ivanov-Smolenskiĭ and Bykov, the official "Lysenkos" of Soviet psychology, physiology, and related disciplines, who in their ruthless advocacy of a raw Pavlovism have exercised great power in these fields since 1950, have recently been censured in *Pravda* along with Lysenko for suppressing the views of those holding conceptions "deviant" from theirs (440).

BIBLIOGRAPHY

(Russian titles are given in English translation in brackets. Many discrepancies will be noted between the listing of items here and as given in Soviet bibliographies. The latter are for a variety of reasons notoriously unreliable: Bogoslovskiĭ's authorship of articles is of late suppressed with substitution of Kravkov (Ed.) in his stead; English titles are given in Russian translation and occasionally in Russian mistranslation; pagination is frequently incorrect or lacking and, since so many Soviet bibliographies are derivative, original errors and omissions remain uncorrected; etc. The number of works on sensory interaction listed herein is undoubtedly impressive, but not as impressive as mere enumeration might suggest. For example, many of the publications represent identical or near-identical articles placed in different journals. In the main, only those works are included in the bibliography which fall within the scope of this paper as indicated in the introduction.)

1. ALEKSEENKO, N. IU. [Binaural effect with excitation of other afferent systems and with influence of an ultra-high-frequency field on the brain.] In *Referaty nauchno-issled. rabot (Otd. biol. nauk).* [Abstracts of research (Division of biological sciences.)] Moscow: Akad. Nauk SSSR, 1947. Pp. 363–364.

2. ALEKSEENKO, N. IU. [Influence of nonacoustic stimulation on the perception of direction of sound.] *Probl.*

fisiol. Akustiki, 1949, 1, 74–88.

3. ALEKSEEV, M. A. [Dynamics of inter-action of the dermal sensory systems in man in microintervals of time.] Uchën. Zap. Leningr. Univer., 1950, No. 123 (Ser. biol. Nauk, No. 22), 400–415.

4. ANDREEV, L. A. Fiziologiia organov chuvstv. [Physiology of the sense organs.] Moscow: Moskovsk. Gos. Univer., 1941.

5. ARKHANGEL'SKIĬ, V. [On the 100th anniversary of Prof. A. V. Ivanov's birth.] Sovetsk. Vestn. Oftal., 1936, 9(4), 570–571.

6. ARVANITAKI, A. Effects evoked in an axon by the activity of a contiguous one. J. Neurophysiol., 1942, 5, 89–108.

7. AVERBAKH, M. I. [From the editor.] Probl. fiziol. Optiki, 1941, 1, 5–6.

8. BABSKIĬ, E. B. [The significance of the sympathetic nervous system in regulation of excitability of the visual analyzer.] Probl. fiziol. Optiki, 1947, 4, 17–30.

9. BABSKIĬ, E. B., & SKULOV, D. K. [On the influence of sympathectomy on dark adaptation of the eye.] Biull. èksper. Biol. i Medits., 1944, 18(1–2), 59–62.

10. BAM, L. A. [Sixth conference on physiological problems.] Fiziol. Zh. SSSR, 1940, 28(6), 707–712.

11. BAM, L. A. [Seventh conference on problems of higher nervous activity, dedicated to the memory of I. P. Pavlov.] Fiziol. Zh. SSSR, 1940, 29(5), 469–475.

12. BARBEL', I. E. [Color sensation in cases of glaucoma.] Vestn. Oftal., 1939, 15 (9–10), 10–22. Also in Vestn. Oftal., 1940, 17(3), 330. (Abstract)

13. BAUER, R. A. The new man in Soviet psychology. Cambridge: Harvard Univer. Press, 1952.

14. BELOSTOTSKIĬ, E. M., & IL'INA, S. A. [The influence of stimulation of the vestibular apparatus on photic sensitivity of the eye.] Vestn. Oftal., 1937, 10, 135–142.

15. BERITOV, I. S. [Account of the activities of the VI All-Union Physiological Congress.] Fiziol. Zh. SSSR, 1937, 23(6), 801–825.

16. BOGOSLOVSKIĬ, A. I. [Conditioned reflex changes of critical flicker frequency in central and peripheral vision.] Sovetsk. Vestn. Oftal., 1936, 8(6), 795–803.

17. BOGOSLOVSKIĬ, A. I. [An experiment in forming sensory conditioned reflexes in man.] Fiziol. Zh. SSSR, 1936, 20(6), 1017–1029.

18. BOGOSLOVSKIĬ, A. I. An attempt at creating sensory conditioned reflexes in humans. J. exp. Psychol., 1937, 21, 403–422.

19. BOGOSLOVSKIĬ, A. I. [On changes in electrical sensitivity of the eye during the course of the day.] Biull. èksper. Biol. i Medits., 1937, 3(2), 140–142.

20. BOGOSLOVSKIĬ, A. I. [On the influence of sound on the electrical sensitivity of the eye.] Biull. èksper. Biol. i Medits., 1937, 3(3), 329–332.

21. BOGOSLOVSKIĬ, A. I. [On the fusion of light flickers, caused by electrical stimulation of the eye.] Biull. èksper. Biol. i Medits., 1937, 3(3), 333–336.

22. BOGOSLOVSKIĬ, A. I. [On the relationship of conditioned and unconditioned sensory reflexes in man.] Vestn. Oftal., 1937, 10(5), 726–734.

23. BOGOSLOVSKIĬ, A. I. [Conditioned reflexes and electrical sensitivity of the eye.] Vestn. Oftal., 1937, 10(6), 896. (Abstract)

24. BOGOSLOVSKIĬ, A. I. Changement de la fréquence critique des papillotements lumineux à caractère de réflexe conditionné. Arch. Ophthal., 1938, 2, 219–227.

25. BOGOSLOVSKIĬ, A. I. [Conditioned reflex modification of the differential sensitivity of the eye to brightness.] Biull. èksper. Biol. i Medits., 1939, 8(3–4), 272–275.

26. BOGOSLOVSKIĬ, A. I. [The influence of dark and light adaptation on the muscular balance of the eyes.] Biull. èksper. Biol. i Medits., 1939, 8(3–4), 276–278.

27. BOGOSLOVSKIĬ, A. I. [The influence of auditory and olfactory stimulation on optic chronaxie.] Biull. èksper. Biol. i Medits., 1939, 8(5), 363–365.

28. BOGOSLOVSKIĬ, A. I. The dependency of the contrast-sensitivity of the eye upon adaptation. Ophthalmologica, 1939, 97, 289–302.

29. BOGOSLOVSKIĬ, A. I. [The influence of visual activity on several sensory functions of the eye in connection with the analysis of visual fatigue.] Fiziol. Zh. SSSR, 1940, 28(4), 292–302.

30. BOGOSLOVSKIĬ, A. I. [Problem of the electrical sensitivity of the eye.] Probl. fiziol. Optiki, 1944, 2, 136–172.

31. BOGOSLOVSKIĬ, A. I. [On the correlation of central and peripheral processes of visual adaptation.] Probl. fiziol. Optiki, 1944, 2, 173–182.

32. Bogoslovskiĭ, A. I., & Kravkov, S. V. [The influence of noise of the airplane motor on vision.] *Probl. fiziol. Optiki*, 1941, 1, 69–75.

33. Bogoslovskiĭ, A. I., Kravkov, S. V., & Semenovskaia, E. N. [The influence of place of preliminary stimulation of the retina by light on subsequent photic and electrical sensitivity of the eye.] *Fiziol. Zh. SSSR*, 1935, 19(4), 814–825.

34. Brandis, S. A. [Changes in the visual functions accompanying different forms of activity.] *Fiziol. Zh. SSSR*, 1937, 23(2), 202–210.

35. Brandis, S. A. [On the methodology of investigating photic sensitivity of the eye in man.] *Biull. èksper. Biol. i Medits.*, 1938, 5(1), 75–78.

36. Brandis, S. A. [Changes in photic sensitivity of the eye in man in connection with physical activity.] *Biull. èksper. Biol. i Medits.*, 1938, 6(3), 341–343.

37. Brandis, S. A. [Changes in photic sensitivity of the eye in man in connection with mental activity.] *Biull. èksper. Biol. i Medits.*, 1939, 7(5), 395–397.

38. Brandis, S. A. [Some data on changes in photic sensitivity of the eye in man in connection with different types of activity engaged in by him.] *Biull. èksper. Biol. i. Medits.*, 1939, 8(6), 435–438.

39. Brandis, S. A. [Changes in photic sensitivity of the eye in man accompanying different types of activity.] *Vestn. Oftal.*, 1939, 14(1), 130. (Abstract)

40. Brandis, S. A. [Changes in the visual functions accompanying different forms of activity.] *Fiziol. Zh. SSSR*, 1940, 29(5), 424–433.

41. Brandis, S. A., & Gorkin, Z. D. [Modification of photic sensitivity of the eye in connection with action on the skin of radiant energy derived from different parts of the spectrum.] *Biull. èksper. Biol. i Medits.*, 1941, 11(1), 56–59.

42. Bransburg, F. S. [The influence of hue of stimulus on the sensitivity of the dark-adapted eye.] *Biull. èksper. Biol. i Medits.*, 1940, 10(1–2), 62–65.

43. Bronshteĭn, A. I. [On the sensibilizing influence of acoustic stimulation on the auditory organ.] *Biull. èksper. Biol. i Medits.*, 1936, 1(4), 276–279; 2(5), 365–367. Also in *Fiziol. Zh. SSSR*, 1936, 20(6), 1051–1061.

44. Bronshteĭn, A. I. [On sensibilization in the determination of sensitivity of the sense organs.] *Fiziol. Zh. SSSR*, 1938, 25(5), 754. (Abstract)

45. Bronshteĭn, A. I. [On sensibilization in the determination of thresholds of sensitivity of sense organs.] *Fiziol. Zh. SSSR*, 1939, 26(6), 587–595.

46. Bronshteĭn, A. I. [On sensibilization to color stimuli.] *Vestn. Oftal.*, 1941, 18(5), 561. (Abstract)

47. Bronshteĭn, A. I. *Sensibilizatsiia organov chuvstv.* [Sensibilization of the sense organs.] Leningrad: Kirov War-Military Academy, 1946.

48. Bronshteĭn, A. I. *Vkus i obonianie.* [Taste and smell.] Moscow: Akad. Nauk SSSR, 1950.

49. Bronshteĭn, A. I., & Lebedinskiĭ, A. V. [On the detection of interactional phenomena between different elements of the retina.] *Fiziol. Zh. SSSR*, 1939, 26(6), 596–602.

50. Bronshteĭn, A. I., & Mil'shteĭn, G. I. [Measurement of conditioned differential thresholds as a method of investigating functional connections of the visual apparatus.] *Probl. fiziol. Optiki*, 1948, 6, 112–120.

51. Bronshteĭn, A. I., & Mil'shteĭn, G. I. [The influence of various factors on conditioned differential thresholds of the visual and tactile analyzers.] *Fiziol. Zh. SSSR*, 1949, 35(2), 154–166.

52. Bronshteĭn, A. I., & Mil'shteĭn, G. I. [An investigation of the functional lability of the visual analyzer utilizing the measurement of conditioned differential thresholds of adequate stimuli.] *Fiziol. Zh. SSSR*, 1950, 36(3), 304–311.

53. Bronshteĭn, A. I., & Zimkin, N. V. [Influence of stimulation of the macular and peripheral fields of the retina on retinal photic sensitivity in determinations utilizing photic stimuli of short duration.] *Fiziol. Zh. SSSR*, 1938, 25(5), 758. (Abstract)

54. Bronshteĭn, A. I., & Zimkin, N. V. [Phasic modifications of peripheral sensitivity of the retina in the process of interaction of its elements.] *Tr. Voenno-Medits. Akad. im. Kirova*, 1941, 34, 194–198.

55. Brücke, E. T. L. A. Orbeli's Untersuchungen über die sympatische Innervation nicht vegetativer Organe. *Klin. Wochensch.*, 1927, 6, 703–704.

56. Bushmich, D. G. [On the influence of biogenous stimuli on functions of the normal eye.] *Vestn. Oftal.*, 1946, 25(6), 46. (Abstract)

57. BYKOV, K. M. *Kora golovnogo mozga i vnutrennie organy.* [The cortex of the brain and the internal organs.] Moscow: Medgiz, 1947.

58. BYKOV, K. M., & RAZENKOV, I. P. (Eds.) *Materialy po fisiologii retseptorov.* [Data on the physiology of the receptors.] Moscow: Medgiz, 1948.

59. CHIRKOVSKIĬ, V. V. [Fifty years of activity of the Leningrad Ophthalmological Society.] *Vestn. Oftal.,* 1949, 28(3), 10–13.

60. CHUBUKOV, A. V. [On erroneous views in several articles of the journal "Herald of Ophthalmology."] *Vestn. Oftal.,* 1950, 29(5), 40–44.

61. CHUPRAKOV, A. T. [Differential sensitivity of the foveal region of the retina as affected by the presence of dark objects in other regions of the visual field.] *Vestn. Oftal.,* 1940, 17(5), 680–685.

62. DAVYDOV, V. G. [The influence, on photic sensitivity of peripheral vision, of contralateral irradiation of dorsal skin with ultraviolet light.] *Vestn. Oftal.,* 1940, 16(6), 537. (Abstract)

63. DAVYDOV, V. G. [The influence, on photic sensitivity of peripheral vision, of unilateral irradiation of the dermal surface with ultraviolet light.] *Probl. fiziol. Optiki,* 1941, 1, 81–86.

64. DEMIRCHOGLIAN, G. G. [Review of S. V. Kravkov's *Color vision.*] *Sovetsk. Kniga,* 1952, No. 4, 39–42.

65. DIONESOV, S. M., & LEBEDINSKIĬ, A. V. [On the dynamics of the coördinated act in the sensory sphere.] *Fiziol. Zh. SSSR,* 1938, 25(5), 758. (Abstract)

66. DIONESOV, S. M., LEBEDINSKIĬ, A. V., & TURTSAEV, IA. P. [On the influence of reflex (cold) stimuli on the sensitivity of the dark-adapted eye.] *Fiziol. Zh. SSSR,* 1934, 17(1), 23–31.

67. DIONESOV, S. M., LEBEDINSKIĬ A. V., TURTSAEV, IA. P., & ZAGORUL'KO, L. T. [The influence of physical effort on dark adaptation of the eye.] *Fiziol. Zh. SSSR,* 1933, 16(5), 733–739.

68. DIONESOV, S. M., LEBEDINSKIĬ, A. V., & ZAGORUL'KO, L. T. [On the interrelation between central and peripheral vision.] *Fiziol. Zh. SSSR,* 1934, 17(3), 560–576.

69. DIONESOV, S. M., LEBEDINSKIĬ, A. V., & ZAGORUL'KO, L. T. [Data contributing to theory on interrelations of the afferent systems.] *Fiziol. Zh. SSSR,* 1936, 21(5–6), 917–918.

70. DIONESOV, S. M., LEBEDINSKIĬ, A. V., & ZAGORUL'KO, L. T. [On the dynamics of the coördinated act in the sensory sphere.] *Fiziol. Zh. SSSR,* 1937, 23(6), 627–635.

71. DOBRIAKOVA, O. A. [A study in the field of electrical sensitivity of the visual and gustatory receptors.] *Biull. eksper. Biol. i Medits.,* 1938, 6(3), 344–347.

72. DOBRIAKOVA, O. A. [On parallelism in modifications of electrical sensitivity of the visual and gustatory organs under the influence of visual and gustatory stimuli.] *Fiziol. Zh. SSSR,* 1939, 26(2–3), 192–199.

73. DOBRIAKOVA, O. A. [The influence of the center of the retina on its periphery.] *Biull. eksper. Biol. i Medits.,* 1941, 11(2), 162–163.

74. DOBRIAKOVA, O. A. [On the influence of gustatory and thermal stimuli on color vision.] In *Tr. 1-ĭ sessii Mosk. ob-va fiziologov, biokhimikov i farmakologov.* [Transactions of the first session of the Moscow Society of Physiologists, Biochemists, and Pharmacologists.] Moscow: Medgiz, 1941. Pp. 80–82.

75. DOBRIAKOVA, O. A. [On changes in photic sensitivity of the retinal periphery as a function of central stimulation in the normal and pathological case.] *Vestn. Oftal.,* 1941, 18(5), 561–562. (Abstract)

76. DOBRIAKOVA, O. A. [The influence of gustatory, warm and cold, and aural stimuli on critical flicker frequency of monochromatic lights.] *Probl. fiziol. Optiki,* 1944, 2, 81–84.

77. DOBRIAKOVA, O. A. [On the sensibilizing action of photic stimulation on the glaucomatous eye.] *Probl. fiziol. Optiki,* 1946, 3, 102–105.

78. DOBRIAKOVA, O. A. [On simultaneous modification of sensitivity of sense organs upon stimulation of one of them.] *Izv. Akad. pedag. Nauk RSFSR,* 1947, No. 8, 33–36.

79. DOBRIAKOVA, O. A. [Formation of a situationally conditioned sensory reflex to one's place of work.] *Izv. Akad. pedag. Nauk RSFSR,* 1947, No. 8, 100–104.

80. DOBRIAKOVA, O. A. [On several possibilities of conditioned-reflex modification of visual sensitivity.] *Probl. fiziol. Optiki,* 1948, 6, 308–314.

81. DOLIN, A. I. [New data contributing to the physiological understanding of association in man (the photochemical conditioned reflex of the eye).] *Arkh.*

biol. Nauk, 1936, 42 (1–2), 275–284, 301.

82. DUBINSKAIA, A. A. [On some factors modifying the sensitivity of night vision.] *Akad. pedag. Nauk RSFSR*, 1947, No. 8, 42–46.

83. DUBINSKAIA, A. A. [Conditioned sensory (visual) reflexes.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 95–97.

84. DUBINSKAIA, A. A. [The influence of the idea of light and darkness on the sensitivity of night vision.] *Izv. Akad. pedag. RSFSR*, 1947, No. 8, 104–107.

85. DZIDZISHVILI, N. N. [The influence of illumination of one eye on several functions of the other.] *Vestn. Oftal.*, 1937, 10(2), 322–325. Also in *Vestn. Oftal.*, 1937, 10(4), 623. (Abstract)

86. DZIDZISHVILI, N. N. [On the influence of intensity of photic stimulation of one eye on the differential sensitivity of the other.] In *Issledovaniia po psikhofiziologii zreniia.* [*Studies on the psychophysiology of vision.*] Moscow: Sotsėkgiz, 1937.

87. DZIDZISHVILI, N. N. [On the variability of the influence of illumination of one eye on the effect of irradiation in the other in the course of repeated experiments.] In *Issledovaniia po psikhofiziologii zreniia.* [Studies on the psychophysiology of vision.] Moscow: Sotsėkgiz, 1937.

88. DZIDZISHVILI, N. N. [On the influence, on differential sensitivity of one eye, of intensity of simultaneous photic stimulation of the other.] *Probl. fiziol. Optiki*, 1941, 1, 43–46.

89. EFIMOV, V. V. [The influence of "imagined" physical effort on the excitability of the visual centers.] *Biull. ėksper. Biol. i Medits.*, 1936, 2(2), 115–117.

90. EFIMOV, V. V. [A new method of dark adaptation of the eye employing blue illumination.] *Biull. ėksper. Biol. i Medits.*, 1936, 2(2), 125–126.

91. EFIMOV, V. V. [On the various effects of large and small muscular activity on the excitability of the nervous centers in man.] *Biull. ėksper. Biol. i Medits.*, 1936, 2(5), 363–364.

92. EFIMOV, V. V. [Review of P. P. Lazarev's *Studies on adaptation*, 1947.] *Sovetsk. Kniga*, 1948, No. 6, 33–37.

93. EFIMOV, V. V., & KAZIMIROVA, S. I. [The exciting and depressing action of different colors of the solar spectrum on the sensitivity of peripheral vision.]

Biull. ėksper. Biol. i Medits., 1936, 2(2), 123–124.

94. EFIMOV, V. V., & VEREHUTINA, A. I. [The influence of rhythm of breathing on night vision and electrical sensitivity of the eye in man.] *Biull. ėksper. Biol. i Medits.*, 1946, 22(3), 54–56.

95. FADEEVA, A. A. [The role of sympathetic innervation in the regulation of the interrelational processes of the afferent systems.] In *Tezisy 11-go soveshch. po fiziol. problemam.* [Theses of the eleventh conference on physiological problems.] Moscow: Akad. Nauk SSSR, 1946.

96. FARFEL', M. N. [The influence of adequate stimulation of the vestibular apparatus on the state of the visual apparatus.] *Biull. ėksper. Biol. i Medits.*, 1937, 4(1), 47–50.

97. FEDOROV, N. T. [On several questions in the theory and practice of adaptation theory.] *Arkh. biol. Nauk*, 1938, 49(1), 152–168.

98. FEDOROV, N. T. [On some general principles concerning the action of accessory stimulation on the differential sensitivity of the eye.] *Fiziol. Zh. SSSR*, 1938, 25(5), 761. (Abstract)

99. FEDOROV, N. T. Some regularities underlying the effect of indirect stimuli (Nebenreiz) upon the discriminatory sensitivity of the eye. *Dokl. Akad. Nauk SSSR*, 1939, 22(2), 70–74.

100. FEDOROV, N. T. [New developments in the theory of vision.] *Priroda*, 1940, 29(2), 42–53.

101. FEDOROV, N. T. [Contemporary state of the problem of color vision.] *Probl. fiziol. Optiki*, 1948, 6, 69–70.

102. FEDOROV, N. T., & FEDOROVA, V. I. Study of colour vision. *Izv. Akad. Nauk SSSR*, 1935, VII Seriia, No. 10, 1431–1450.

103. FEDOROV, N. T., & FEDOROVA, V. I. [On the curve of spectral sensitivity of the eye.] *Biull. ėksper. Biol. i Medits.*, 1936, 2(1), 52–54.

104. FEDOROV, N. T., & FEDOROVA, V. I. On the problem of the curve of spectral sensitivity of the eye. *Dokl. Akad. Nauk SSSR*, 1936, 11(9), 377–380.

105. FEDOROV, N. T., & FEDOROVA, V. I. [On the influence of a glaring light source on the size of the visual field for different colors.] *Probl. fiziol. Optiki*, 1946, 3, 109–112.

106. FEDOROV, N. T., & KUZNETSOV, A. I. [The action of sympathomimetic amines on the electrical sensitivity of

the eye.] *Probl. fiziol. Optiki*, 1949, 7, 39–46.

107. FEDOROVA, V. I. [On the eye's sensitivity to change of hue.] *Zh. prikl. Fiziki*, 1928, 5(3), 4–10.

108. FILATOV, V. P., & KASHUK, M. È. [The influence of muscular activity on the acuity of vision for normal eyes.] *Oftal. Zh.*, 1946, 1(4), 3–10.

109. FREIMAN, S. IA. [Evaluation of the first Soviet edition of Ishihara tables as compared with the Japanese in natural and artificial light.] *Vestn. Oftal.*, 1938, 12(5–6), 586–593.

110. FROLOV, IU. P. [Summary of the joint session of the USSR Academy of Sciences, the USSR Academy of Medical Sciences, and the All-Union Society of Physiologists . . .] *Biull. èksper. Biol. i Medits.*, 1946, 21 (4), 75–77.

111. GALOCHKINA, L. P. [Inductive processes in the visual apparatus in their dependence on color of stimuli, anomalies of color vision, and certain pharmacological influences.] *Probl. fiziol. Optiki*, 1941, 1, 17–24.

112. GALOCHKINA, L. P. [On the influence on vision (of the berries) of the Chinese little-lemon bush (*Schizandra Chinensis*).] *Vestn. Oftal.*, 1945, 24(5–6), 86. (Abstract)

113. GALOCHKINA, L. P. [The processes of inductive modification of sensitivity for different parts of the spectrum.] *Probl. fiziol. Optiki*, 1946, 3, 80–93.

114. GALOCHKINA L. P. [On the action (of the berries) of the Chinese little-lemon bush, *Schizandra Chinensis*, on the photic and chromatic sensitivity of the eye.] *Probl. fiziol. Optiki*, 1948, 5, 71–73.

115. GAPEEV, P. I., & ROMANOVA-BOKHON, O. A. [The influence of phenamine on several visual functions.] *Vestn. Oftal.*, 1948, 27(1), 36–40.

116. GASSOVSKIĬ, L. N. [On the reorganization of the production of eye lenses and ophthalmological instruments in the USSR.] *Sovetsk. Vestn. Oftal.*, 1934, 5(6), 583. (Abstract)

117. GASSOVSKIĬ, L. N. [Present appliances for eye correction and their manufacture in the USSR.] *Vestn. Oftal.*, 1935, 6(4), 583. (Abstract)

118. GASSOVSKIĬ, L. N. [Illumination of the nonobserving eye in work with monocular optical instruments.] *Probl. fiziol. Optiki*, 1941, 1, 33–42.

119. GASSOVSKIĬ, L. N., & KHOKHLOVA, A. N. [The influence of illumination of one eye on the adaptive process in the other.] *Tr. In-ta Tochn. Mekhan. i Optiki*, 1939, 1(3).

120. GELLER, I. M. [The influence of nation of one eye on the photic and ... *Vopr. aviats. fiziol.* [Problems of aviation physiology.] Moscow: Medgiz, 1938. Pp. 171–183.

121. GERARD, R. W., MARSHALL, W. H., & SAUL, L. J. Electrical activity of the cat's brain. *Arch. Neurol. Psychiat.*, 1936, 36, 675–738.

122. GERSHUNI, G. V. [Conference on the physiology of the sense organs.] *Priroda*, 1938, 27(9), 92–98.

123. GERSHUNI, G. V. [Physiology of the sense organs.] In L. A. Orbeli (Ed.), *Uspekhi biologicheskikh nauk v SSSR za 25 let.* [Progress of the biological sciences in the USSR after 25 years.] Moscow: Akad. Nauk SSSR, 1945. Pp. 58–64.

124. GERSHUNI, G. V. [On the interrelationships between sensation and the conditioned reflex.] *Fiziol. Zh. SSSR*, 1946, 32(1), 43–48.

125. GERSHUNI, G. V. [A study of subsensory reactions during sense-organ activity.] *Fiziol. Zh. SSSR*, 1947, 33(4), 393–412.

126. GERSHUNI, G. V. [On unsensed reactions (subsensory activity) upon stimulation of the sense organs.] In K. M. Bykov (Ed.), *Materialy po fiziologii retseptorov.* [Data on the physiology of the receptors.] Moscow: Medgiz, 1948. Pp. 23–29.

127. GERSHUNI, G. V. [Reflex reactions in their connection with sensations in the course of applying external stimulation to the sense organs of man.] *Fiziol. Zh. SSSR*, 1949, 35(5), 541–560.

128. GERSHUNI, G. V. [On reorganization of the auditory function under the action of sound.] *Probl. fiziol. Akustiki*, 1949, 1, 5–20.

129. GERSHUNI, G. V. [On a quantitative study of the limits of action of acoustic stimulation.] *Probl. fiziol. Akustiki*, 1950, 2, 29–36.

130. GOL'DBURT, S. N. [The significance of macro- and micro-intervals between stimulations in research on the interaction of the sense organs.] In *Chetvert. soveshch. po fiziol. probl.; fiziol. organov chuvstv.* [Fourth conference on physiological problems; physiology of the sense organs.] Moscow: Akad. Nauk SSSR, 1938. Pp. 15–16. Also in *Fiziol. Zh. SSSR*, 1938, 25(5), 755. (Abstract)

micro- and macro-in... ....
....
action of the organs of vision and
hearing.] *Arkh. biol. Nauk*, 1940,
60(1), 24–32.

GOL'DBURT, S. N. [On the interaction of
vision and hearing in microintervals of
time.] *Biull. èksper. Biol. i Medits.*,
1946, 22(3), 36–39.

GOL'DBURT, S. N. [On the emergence of
auditory sensations upon monocular
electrical stimulation for fixed inter-
vals of time.] *Biull. èksper. Biol. i
Medits.*, 1946, 22(4), 27–30.

GOL'TS, E. P., & SHEIVEKHMAN, B. E.
[The influence of adequate stimulation
of the olfactory receptor on change of
auditory threshold.] In K. M. Bykov
(Ed.), *Materialy po fiziologii retsep-
torov.* [Data on the physiology of the
receptors.] Moscow: Medgiz, 1948.
Pp. 30–34.

135. GORKIN, Z. D., & BRANDIS, S. A. [On
photic sensitivity of the eye as an in-
dex of the adaptation of the organism
to ultraviolet irradiation.] *Biull.
èksper. Biol. i Medits.*, 1939, 8(5),
366–368.

136. GUREVICH, M. M. *TSvet i ego izmerenie.*
[Color and its measurement.] Mos-
cow: Akad. Nauk SSSR, 1950.

137. GURTOVOĬ, G. K. [On the direction of
changes in absolute photic sensitivity
of the eye in twilight vision during the
action of an aural stimulus.] *Biull.
èksper. Biol. i Medits.*, 1947, 24(1),
50–53.

138. GURTOVOĬ, G. K. [Dark adaptation in
achromats.] *Dokl. Akad. Nauk SSSR*,
1947, 58(8), 1845–1848.

139. GURTOVOĬ, G. K. *Svetovaia i èlektriche-
skaia chuvstvitel'nost' glaza akhromatov
i vliianie na neë nepriamykh razdra-
zhiteleĭ.* [Photic and electrical sensi-
tivity of the achromatic eye and the
influence on it of indirect stimuli.]
Moscow: Akad. Nauk SSSR, 1947.

140. GURTOVOĬ, G. K. [An analysis of the
role of the cone-apparatus on the basis
of experiments with achromats.]
Probl. fiziol. Optiki, 1949, 7, 94–110.

141. GURTOVOĬ, G. K. [Reactions to color in
the color-blind (achromats).] *Probl.
fiziol. Optiki*, 1950, 9, 90–105.

142. GURTOVOĬ, G. K. *Svoĭstva zreniia akhro-
matov.* [Properties of vision in achro-
mats.] Moscow: Akad. Med. Nauk
SSSR, 1950.

143. GURTOVOĬ, G. K., & KRAVKOV, S. V.
[The influence of hue of illumination of

... ...
... ...
7a... ... ...

144. IVANOV..., ... A. [... ... ...
on the magnitude of the visual field.]
I... S. V. ... ... M. ... ...
(Eds.), *Zritel'nye oshchushcheniia i
... ...
... ... M. ... ...
... ...
... ...
Also in *Tr. 1-1 konfer. po fiziol. optike*,
1946, 232–235.

145. IAKOVLEV, P. A. The influence of acous-
tic stimuli upon the limits of visual
fields for different colors. *J. opt. Soc.
Amer.*, 1938, 28, 286–289.

146. IAKOVLEV, P. A. [The influence of
acoustic ... ...
the ... ...
... ... 17... ... ...

147. IA... ... ... S. P. ... [The influence of ...
stillation of pilocarpine on Rayleigh's
equation.] *Probl. fiziol. Optiki*, 1941, 16...
361–362.

148. IAKOVLEVA, S. P. [The influence of in-
stillation of pilocarpine on Rayleigh's
equation.] *Probl. fiziol. Optiki*, 1941,
1. 97–98.

149. IARMOLENKO, A. V. [Development of
distant ... ...]
*Uchën. Zap. Leningr. Univer. (Ser. filos.
Nauk)*, 1948, No. 2, 195–207.

150. IUDINA, E. F. [Interaction between the
olfactory and the vestibular appara-
tus. The influence of geraniol on the
excitability of the vestibular appara-
tus in man.] *Vestn. Oto-rino-laring.*,
1940, 5(9), 3–8.

151. IUDINA, E. F. [Interrelation between
the olfactory and vestibular appara-
tus. The influence on the vestibular
apparatus of stimulating the olfactory
receptor with thymol.] *Vestn. Oto-
rino-laring.*, 1941, 6(3–4), 32–35.

152. JASPER, H. H., & MONNIER, A. M.
Transmission of excitation between
excised non-myelinated nerves. An
artificial synapse. *J. cell. comp.
Physiol.*, 1938, 11, 259–277.

153. KAL'KUTINA, M. L. [On the influence of
pilocarpine on the photic sensitivity
of the glaucomatous eye.] *Vestn.
Oftal.*, 1941, 18(3), 250–253. Also in
Vestn. Oftal., 1940, 16(6), 533.
(Abstract)

154. KAMINSKIĬ, D. S. [Ophthalmology and
the planning of scientific work.]
Sovetsk. Vestn. Oftal., 1932, 1(5),
175–182.

155. KASHUK, M. E. [Influence of biog-

enous stimulants on normal visual functions.] *Probl. fiziol. Optiki*, 1948, 6, 238–245.

156. KATZ, B., & SCHMITT, O. H. Electrical interaction between two adjacent nerve fibers. *J. Physiol.*, 1940, 97, 471–488.

157. KEKCHEEV, K. KH. [On the influence of inadequate stimuli on the sensitivity of peripheral vision. Report No. IV: Measurement of the speed of the first phase of the period of dark adaptation.] *Biull. èksper. Biol. i Medits.*, 1937, 4(4), 341–342.

158. KEKCHEEV, K. KH. On the action of non-adequate stimuli on receptors. *Dokl. Akad. Nauk SSSR.*, 1937, 14(8), 495–497.

159. KEKCHEEV, K. KH. [On the influence of inadequate stimuli on the sensitivity of achromatic vision. Report No. V: The influence of muscular activity.] *Biull. èksper. Biol. i Medits.*, 1938, 5(4), 432–436.

160. KEKCHEEV, K. KH. [On the so-called vegetative reflexes.] *Klinich. Medits.*, 1940, 18(4), 35–39.

161. KEKCHEEV, K. KH. [On the influence of excitation of other sense organs on the visual functions.] *Vestn. Oftal.*, 1940, 16(4), 246–253.

162. KEKCHEEV, K. KH. [On the modification of thresholds of achromatic vision in man through action of ultrashort, ultraviolet, and X-rays.] *Probl. fiziol. Optiki*, 1941, 1, 77–79.

163. KEKCHEEV, K. KH. Mechanism of sensitivity changes of sense organs. *Nature.*, 1942, 150, No. 3808, 491–492.

164. KEKCHEEV, K. KH. [Psychophysiology of camouflage and reconnaissance.] Moscow: Sovetsk. Nauka, 1942.

165. KEKCHEEV, K. KH. Expediting visual adaptation to darkness. *Nature*, 1943, 151, No. 3839, 617–618.

166. KEKCHEEV, K. KH., *et al.* Problem of night vision. *War Med.*, 1943, 3(2), 171–173.

167. KEKCHEEV, K. KH. The problem of night vision. *Amer. Rev. Soviet Med.*, 1944, 1(4), 300–302.

168. KEKCHEEV, K. KH. *Psychophysiology of camouflage and reconnaissance.* London: 1944. (Publisher not known.)

169. KEKCHEEV, K. KH. Conditioned excitors and human sense organs. *Nature*, 1945, 156, No. 3967, 573–574.

170. KEKCHEEV, K. KH. [Participation of central factors in the activity of the eye.] *Vestn. Oftal.*, 1945, 24(5–6), 85–86. (Abstract)

171. KEKCHEEV, K. KH. [Conditioned sensory connections.] *Fiziol. Zh. SSSR*, 1946, 32(2), 159–162.

172. KEKCHEEV, K. KH. *Nochnoe zrenie (kak luchshe videt' v temnote).* [Night vision (how better to see in the dark).] Moscow: Sovetsk. Nauka, 1946.

173. KEKCHEEV, K. KH. [On reflex modification of adaptive-trophic influence of the autonomic nervous system on excited tissues of the human organism.] *Fiziol. Zh. SSSR*, 1947, 33(4), 475–482.

174. KEKCHEEV, K. KH. [On the ways of investigating efficiency in man.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 3–26.

175. KEKCHEEV, K. KH. [The problem of physical and mental efficiency in the light of contemporary ideas.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 115–138.

176. KEKCHEEV, K. KH. Some facts concerning the problem of mental efficiency. *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 139–141.

177. KEKCHEEV, K. KH. [Role of central factors in the sensitivity of the eye.] *Probl. fiziol. Optiki*, 1948, 6, 291–299.

178. KEKCHEEV, K. KH., *et al.* [On change in thresholds of achromatic vision under influence of excitation of other sense organs.] *Vestn. Oftal.*, 1940, 16(6), 536. (Abstract)

179. KEKCHEEV, K. KH., ANISIMOVA, A. P., & SHAEVICH, Z. Z. [On the influence of inadequate stimuli on sensitivity of achromatic vision. Report No. VII: The action of X-rays.] *Biull. èksper. Biol. i Medits.*, 1939, 7(5), 391–392.

180. KEKCHEEV, K. KH., & DUBINSKAIA, A. A. [The effect of rhythmic acoustic stimulation on the sensitivity of night vision.] *Probl. fiziol. Optiki*, 1946, 3, 106–108.

181. KEKCHEEV, K. KH., KAVTORINA, A. V. & SHLIAPNIKOVA, O. A. [Inadequate action of stimuli on sensitivity of achromatic vision. Report No. VIII: Summation of effects with the action of two stimuli.] *Biull. èksper. Biol. i Medits.*, 1940, 10(3), 187–190.

182. KEKCHEEV, K. KH., KRAVKOV, S. V., & SHVARTS, L. A. [On factors reducing the activity of the visual and auditory organs.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 46–51.

183. KEKCHEEV, K. KH., & MATIUSHENKO, O. A. [The influence of sensory stimulation on sensitivity of peripheral vision.] *Biull. èksper. Biol. i Medits.*, 1936, 2(5), 358–360.

184. KEKCHEEV, K. KH., MATIUSHENKO, O. A., & ORLIUK, A. G. [The influence of inadequate stimuli on sensitivity of peripheral vision. Report No. I: The influence of dosage (intensity and duration of action) of the stimulus.] *Arkh. biol. Nauk*, 1937, 45(3), 157–162.

185. KEKCHEEV, K. KH., MERTYCHEVA, L. I., & SYROVATKO, F. A. [On changes in thresholds of achromatic vision under the influence of enteroceptive impulses and muscular activity.] *Fisiol. Zh. SSSR*, 1938, 25(5), 761. (Abstract)

186. KEKCHEEV, K. KH., & ORLIUK, A. G. [The influence of sensory (gustatory) stimulation on sensitivity of peripheral vision.] *Biull. èksper. Biol. i Medits.*, 1936, 2(5), 361–362.

187. KEKCHEEV, K. KH., & ORLUIK, A. G. [The influence of inadequate stimulation on sensitivity of peripheral vision. Report No. III. The influence of ultraviolet radiation.] *Biull. èksper. Biol. i Medits.*, 1937, 3(4), 353–354.

188. KEKCHEEV, K. KH., & OSTROVSKIĬ, E. P. [On the detection of atmospheric oscillations of ultrasonic frequency by means of measurement of visual thresholds.] *Dokl. Akad. Nauk SSSR*, 1941, 31(4).

189. KEKCHEEV, K. KH., & SHVARTS, L. A. [The sensitivity of twilight vision in the drowsy state.] *Probl. fisiol. Optiki*, 1946, 3, 123–125.

190. KEKCHEEV, K. KH., & SYROVATKO, F. A. [On the problem of enteroceptive stimulation.] *Akush. i Ginekol.*, 1939, 4(5), 17–21.

191. KEKCHEEV, K. KH., & SYROVATKO, F. A. [On the influence of inadequate stimulation on sensitivity of achromatic vision. Report No. VI. The influence of enteroceptive stimulation.] *Biull. èksper. Biol. i Medits.*, 1939, 7(4), 320–323.

192. KHARITONOV, S. A. [On the theory of interaction of the afferent systems.] *Biull. èksper. Biol. i Medits.*, 1936, 1(3), 211–212.

193. KHARITONOV, S. A. [On the interaction of afferent systems.] *Fisiol. Zh. SSSR*, 1938, 25(5), 754. (Abstract)

194. KHARITONOV, S. A. [On the physiological characteristics of the gustatory analyzer.] *Fisiol. Zh. SSSR*, 1940, 29(5), 472. (Abstract)

195. KHARITONOV, S. A. [Interaction of the afferent systems of the gustatory apparatus.] In *Tr. 1-ĭ sessii Mosk. ob-va fisiologov, biokhimikov, i farmakologov.* [Transactions of the first session of the Moscow Society of Physi-ologists, Biochemists, and Pharmacologists.] Moscow: Medgiz, 1941. Pp. 271–272.

196. KHARITONOV, S. A., & ANISIMOVA, A. P. [The dynamics of changes in threshold sensitivity of peripheral vision in cases of traumatic injuries of the nervous system.] *Probl. fisiol. Optiki*, 1948, 6, 245–252.

197. KLASS, IU. A., & CHISTOVICH, L. A. [On the influence of inaudible acoustic stimulation under conditions of binaural interaction.] *Probl. fisiol. Akustiki*, 1950, 2, 37–50.

198. KLIMOVICH, E. F., & MEN'SHUTIN, M. A. [Tests for determination of visual acuity.] *Vestn. Oftal.*, 1938, 13(3), 382–387.

199. KOLBANOVSKIĬ, V. [Review of S. V. Kravkov's *Essay on the general psychophysiology of the sense organs*, 1946.] *Sovetsk. Kniga*, 1946, No. 12, 34–36.

200. KOLEN, A. A. In *Voprosy fisiologii i patologii zreniia.* [Problems in the physiology and pathology of vision.] Moscow: Medgiz, 1950.

201. KOLEN, A. A. [Significance of Pavlov's theories for ophthalmology.] *Vestn. Oftal.*, 1951, 30(1), 5–10.

202. KOROTKIN, I. I. [On change in perception of pitch of tones in connection with several cortical processes.] In *Deviatoe sosveshchanie po fisiol. problemam. Tezisy.* [Ninth conference on physiological problems. Theses.] Moscow: Akad. Nauk SSSR, 1941.

203. KRAVKOV, S. V. [On the course of extinction of afterimages in central vision.] *Zh. prikl. Fisiki*, 1924, 1(1–4).

204. KRAVKOV, S. V. [On the extinction of afterimages resulting from colored excitations.] *Zh. prikl. Fisiki*, 1924, 1(1–4).

205. KRAVKOV, S. V. [On the absorption of light in the yellow spot of the living eye.] *Zh. prikl. Fisiki*, 1925, 2(1–2).

206. KRAVKOV, S. V. [Distribution of brightness in the spectrum and the course of dark adaptation in people with inherited complete color blindness.] *Zh. prikl. Fisiki*, 1927, 4(2), 89–95.

207. KRAVKOV, S. V. [On adaptation of the eye to color stimuli.] *Zh. prikl. Fisiki*, 1928, 5(2), 99–116.

208. KRAVKOV, S. V. [On the apparent difference of complementary colors and colors of successive contrast.] *Zh. prikl. Fisiki*, 1928, 5(suppl.), 115–125. Also in *Izv. Inst. biol. Fisiki*, 1929, 4(2).

209. KRAVKOV, S. V. [The acuity of vision of one eye as a function of illumination

of the other.] *Zh. prikl. Fiziki*, 1930, 7(4), 99–108.

210. KRAVKOV, S. V. [On the dependence of acuity of vision on auditory stimulation.] *Zh. prikl. Fiziki*, 1930, 7(4), 109–112.

210a. KRAVKOV, S. V. Die Beeinflussung der Sehschärfe vom Gehirn aus. *Umschau*, 1930, No. 52.

211. KRAVKOV, S. V. Ueber eine zentrale Beeinflussung der Sehschärfe. *Arch. Ophthal.*, 1930, 124(1), 76–86.

212. KRAVKOV, S. V. Ueber die Abhängigkeit der Sehschärfe vom Schallreiz. *Arch. Ophthal.*, 1930, 124(2), 334–338.

213. KRAVKOV, S. V. Ueber die Beeinflussung der Unterschiedsempfindlichkeit des Auges durch Nebenreize. *Arch. Ophthal.*, 1932, 128(1), 105–111.

214. KRAVKOV, S. V. Der Lichtirradiationseffekt im Auge in seiner Abhängigkeit von den Gesichts- Gehörs- und Geruchsnebenreizen. *Arch. Ophthal.*, 1933, 129(3), 440–451.

215. KRAVKOV, S. V. Die Abhängigkeit des Irradiationseffektes im Auge von der Lichtintensität, Konstrast und Nebenreizwirkung. *Arch. Ophthal.*, 1934, 132(4), 379–398.

216. KRAVKOV, S. V. Die Unterschiedsempfindlichkeit eines Auges unter dem Einfluss vom Schall oder Beleuchtung des anderen Auges. *Arch. Ophthal.*, 1934, 132(4), 421–429.

217. KRAVKOV, S. V. [Influence of sound and illumination of one eye on contrast-sensitivity of the other.] *Arkh. Oftal.*, 1934, No. 172.

218. KRAVKOV, S. V. Changes of visual acuity in one eye under the influence of the illumination of the other or of acoustic stimuli. *J. exp. Psychol.*, 1934, 17, 805–812.

219. KRAVKOV, S. V. Actions des excitations auditives sur la fréquence critique des papillotements lumineux. *Acta Ophthal.*, 1935, 13, 260–272.

220. KRAVKOV, S. V. [On the influence of aural stimulation on flicker fusion.] *Fisiol. Zh. SSSR*, 1935, 19(4), 826–833.

221. KRAVKOV, S. V. [The influence of accessory stimulation on the functions of vision.] *Sovetsk. Vestn. Oftal.*, 1935, 6(4), 583. (Abstract)

222. KRAVKOV, S. V. [On the influence of accessory stimulation on the visual function.] In S. V. Kravkov & B. M. Teplov (Eds.), *Zritel'nye oshchushcheniia i vospriiatiia*. [Visual sensation and perception.] Moscow: Sotsékgiz, 1935. Pp. 86–110.

223. KRAVKOV, S. V. The influence of sound upon the light and colour sensitivity of the eye. *Acta Ophthal.*, 1936, 14, 348–360.

224. KRAVKOV, S. V. [The influence of accessory stimulation on the visual functions.] *Fisiol. Zh. SSSR*, 1936, 21(5–6), 916–917. Also in *Tr. 1-i konfer. po fisiol. optike*, 1936, 25–26.

225. KRAVKOV, S. V. [The action of aural stimulation on photic and chromatic sensitivity of the eye.] *Sovetsk. Vestn. Oftal.*, 1936, 8(6), 787–794.

226. KRAVKOV, S. V. Effect of indirect light stimulation as a function of the intensity of a direct stimulus. *Acta Ophthal.*, 1937, 15, 96–103.

227. KRAVKOV, S. V. The influence of acoustic stimulation upon the colour-sensibility of a protanopic eye. *Acta Ophthal.*, 1937, 15, 337–342.

228. KRAVKOV, S. V. [The influence of aural stimulation on photic and chromatic sensitivity of the eye.] *Isv. Akad. Nauk SSSR, Seriia Biol.*, 1937, No. 1, 237–245.

229. KRAVKOV, S. V. [Research in the field of the physiology of the sense organs.] *Usp. sovrem. Biol.*, 1937, 7(3), 381–394.

230. KRAVKOV, S. V. [The influence of aural stimulation on Rayleigh's equation.] *Vestn. Oftal.*, 1937, 10(1), 121–123.

231. KRAVKOV, S. V. [The influence of aural stimulation on the color sensitivity of the protanopic eye.] *Vestn. Oftal.*, 1937, 11(1), 102–105.

232. KRAVKOV, S. V. [The influence of acoustic stimulation on the color sensitivity of the protanope.] *Vestn. Oftal.*, 1937, 11(1), 102–105.

233. KRAVKOV, S. V. [Physiological optics in the USSR after twenty years.] *Vestn. Oftal.*, 1937, 11(4), 468–478.

234. KRAVKOV, S. V. La vision des couleurs et les excitations auditives. *X V Concilium Ophthalmologicum, Égypte. Résumés des communications*, 1937, 147–149.

235. KRAVKOV, S. V. The influence of the dark adaptation on the critical frequency of flicker for monochromatic lights. *Acta Ophthal.*, 1938, 16, 375–384.

236. KRAVKOV, S. V. Illumination and the visual acuity. *Acta Ophthal.*, 1938, 16(2–3), 385–395.

237. KRAVKOV, S. V. [Interaction of the sense organs.] *Novyi Mir*, 1938, 14(11), 162–170.

238. KRAVKOV, S. V. [New developments in the physiology of vision.] *Priroda*, 1938, 27(3), 42–50.

239. KRAVKOV, S. V. [On visual acuity as a

function of illumination.] *Vestn. Oftal.*, 1938, 12(4), 525–531.

240. KRAVKOV, S. V. [The influence of dark adaptation of the critical flicker frequency of monochromatic light.] *Vestn. Oftal.*, 1938, 13(1), 72–77.

241. KRAVKOV, S. V. [Problems of the physiology of vision being studied in the laboratory of physiological optics.] *Vestn. Oftal.*, 1938, 13(5), 702–703. (Abstract)

242. KRAVKOV, S. V. Some new findings on colour vision. *Acta Medica, URSS*, 1939, 2(3), 461–471.

243. KRAVKOV, S. V. The influence of caffeine on the color-sensitivity. *Acta Ophthal.*, 1939, 17, 89–94.

244. KRAVKOV, S. V. The influence of the loudness of the indirect sound stimulus on the color sensitivity of the eye. *Acta Ophthal.*, 1939, 17, 324–331.

245. KRAVKOV, S. V. The influence of odours upon colour-vision. *Acta Ophthal.*, 1939, 17, 425–441.

246. KRAVKOV, S. V. The fusion of light flickers and accessory stimuli. *Dokl. Akad. Nauk SSSR*, 1939, 22(2), 64–66.

247. KRAVKOV, S. V. On some correlations of different receptors in our colour vision. *Dokl. Akad. Nauk SSSR*, 1939, 22(2), 67–69.

248. KRAVKOV, S. V. On the interrelation of the receptors of colour vision. *Dokl. Akad. Nauk SSSR*, 1939, 22(2), 88–91.

249. KRAVKOV, S. V. [The influence of caffeine on color sensitivity.] *Vestn. Oftal.*, 1939, 14(6), 61–63.

250. KRAVKOV, S. V. [The influence of loudness of accessory aural stimulus on color sensitivity of the eye.] *Vestn. Oftal.*, 1939, 15(1), 100–104.

251. KRAVKOV, S. V. [The influence of odors on color vision.] *Fiziol. Zh. SSSR*, 1940, 28(4), 313–322.

252. KRAVKOV, S. V. Color vision and the autonomic nervous system. *J. opt. Soc. Amer.*, 1941, 31, 335–337.

253. KRAVKOV, S. V. [On the connections of color vision with the autonomic nervous system.] *Probl. fiziol. Optiki*, 1941, 1, 87–95.

254. KRAVKOV, S. V. [An analysis of the nature of color vision.] *Vestn. Oftal.*, 1941, 18(5), 562. (Abstract)

255. KRAVKOV, S. V. [The effect of strychnine on the differential sensitivity of the eye.] *Dokl. Akad. Nauk SSSR*, 1942, 37(4), 167–170.

256. KRAVKOV, S. V. [Physiological optics in the Soviet Union after twenty-five years.] *Vestn. Oftal.*, 1942, 21(6), 42–55.

257. KRAVKOV, S. V. [An analysis of the action of accessory stimuli on vision.] *Probl. fiziol. Optiks*, 1944, 2, 78–80.

258. KRAVKOV, S. V. [Physiological optics during the Great Patriotic War and its role in combatting war injuries of the visual organ.] *Vestn. Oftal.*, 1945, 24(5–6), 86. (Abstract)

259. KRAVKOV, S. V. [The conference on physiological optics, sponsored by the USSR Academy of Sciences.] *Biull. eksper. Biol. i Medits.*, 1946, 22(2), 74–75.

260. KRAVKOV, S. V. [New data on the nature of color vision.] *Vestn. Oftal.*, 1946, 25(6), 45. (Abstract)

261. KRAVKOV, S. V. *Ocherk obshchei psikhofiziologii organov chuvstv.* [Essay on the general psychophysiology of the sense organs.] Moscow: Akad. Nauk SSSR, 1946.

262. KRAVKOV, S. V. [The change in spectral sensitivity of cone-vision under the influence of potassium ions.] *Dokl. Akad. Nauk SSSR*, 1947, 58(7), 1361–1364.

263. KRAVKOV, S. V. [On some principles of vision as a function of accessory stimuli.] *Probl. fiziol. Optiki*, 1947, 4, 31–45.

264. KRAVKOV, S. V. [On the nature of color vision.] *Probl. fiziol. Optiki*, 1948, 6, 13–30.

265. KRAVKOV, S. V. [On the interaction of the sense organs.] In S. L. Rubinshtein (Ed.), *Issledovanie po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 23–42.

266. KRAVKOV, S. V. [Research on color vision.] In S. L. Rubinshtein (Ed.), *Issledovanie po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 371–372.

267. KRAVKOV, S. V. *Vzaimodeistvie organov chuvstv.* [The interaction of the sense organs.] Moscow: Akad. Nauk SSSR, 1948.

268. KRAVKOV, S. V. [The influence of color adaptation of the eye on its reactivity in relation to inadequate stimuli.] *Dokl. Akad. Nauk SSSR*, 1949, 64(2), 203–206.

269. KRAVKOV, S. V. [Contemporary theory of color vision.] *Nauka i Zhisn'*, 1949, 16(5), 17–20.

270. KRAVKOV, S. V. [K. KH. Kekcheev (necrology).] *Probl. fiziol. Optiki*, 1949, 7, 133–134.

271. KRAVKOV, S. V. [Review of Prof. E. ZH. Tron's book, *The variability of the ele-*

ments of the optic apparatus and its clinical significance.] Vestn. Oftal., 1949, 28(4), 45–46.

272. KRAVKOV, S. V. [Ophthalmology and I. P. Pavlov's conditioned-reflex theory.] Vestn. Oftal., 1949, 28(6), 4–6.

273. KRAVKOV, S. V. Rabota organov chuvstv. [The functions of the sense organs.] Moscow: Pravda, 1949.

274. KRAVKOV, S. V. Glaz i ego rabota. [The eye and its functions.] Moscow: Akad. Nauk SSSR, 1950.

275. KRAVKOV, S. V. TSvetovoe zrenie. [Color vision.] Moscow: Akad. Nauk SSSR, 1951.

276. KRAVKOV, S. V., & BILETSKII, G. S. Die Abhängigkeit des Irradiationseffektes im Auge von der Lichtintensität, Kontrast, und Nebenreizwirkung. Arch. Ophthal., 1934, 132(4), 379–398.

277. KRAVKOV, S. V., & BILETSKII, G. S. [Irradiation of light in the eye as a function of brightness, contrast, and simultaneous auditory stimulation.] In Tr. 1-i konfer. po fiziol. optike. [Transactions of the first conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1936. Pp. 227–229.

278. KRAVKOV, S. V., & BILETSKII, G. S. [Irradiation of light in the eye as a function of brightness, contrast, and simultaneous aural stimulation.] Vestn. Oftal., 1935, 6(4), 583. (Abstract)

279. KRAVKOV, S. V., & GALOCHKINA, L. P. [On the action of phenamine on vision.] Biull. eksper. Biol. i Medits., 1943, 16(3), 6–10.

280. KRAVKOV, S. V., & GALOCHKINA, L. P. [Electrotonus in color vision.] Dokl. Akad. Nauk SSSR, 1945, 48(1), 21–22.

281. KRAVKOV, S. V., & GALOCHKINA, L. P. [The influence of potassium and calcium ions on color vision.] Dokl. Akad. Nauk SSSR, 1946, 51(5), 347–350.

282. KRAVKOV, S. V., & GALOCHKINA, L. P. [Color vision as affected by calcium and potassium ions.] Dokl. Akad. Nauk SSSR, 1946, 51(5), 351–352.

283. KRAVKOV, S. V., & GALOCHKINA, L. P. [The effect of direct current on vision.] Probl. fiziol. Optiki, 1947, 4, 77–86.

284. KRAVKOV, S. V., & GALOCHKINA, L. P. [The influence of ions of potassium and calcium on color sensitivity of the eye.] Probl. fiziol. Optiki, 1948, 5, 65–70.

285. KRAVKOV, S. V., & MUZILEV, F. I. [On the interaction of the retinal center and periphery.] Vestn. Oftal., 1939, 14(1), 131. (Abstract)

286. KRAVKOV, S. V., & NIKOFOROVA, O. I [The heightening of sensitivity of peripheral vision by means of preliminary illumination of the eye with red light.] Biull. eksper. Biol. i Medits. 1941, 11(2), 160–161.

287. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. Steigerung der Lichtempfindlichkeit des Auges durch vorangehende Lichtreize. Arch. Ophthal., 1933, 130(4), 513–526.

288. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [The influence of illumination of one eye on the subsequent photic sensitivity of the other.] In S. V. Kravkov & B. M. Teplov (Eds.), Zritel'nye oshchushcheniia i vospriiatiia. [Visual sensation and perception.] Moscow: Sotsekgiz, 1935. Pp. 138–141.

289. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [Influence of illumination of one eye on the photic and electrical sensitivity of the other.] In S. V. Kravkov, & B. M. Teplov (Eds.), Zritel'nye oshchushcheniia i vospriiatiia. [Visual sensation and perception.] Moscow: Sotsekgiz, 1935. Pp. 167–170.

290. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [On the influence of prolonged hunger on the functions of vision.] In S. V. Kravkov & B. M. Teplov (Eds.), Zritel'nye oshchushcheniia i vospriiatiia. [Visual sensation and perception.] Moscow: Sotsekgiz, 1935. Pp. 177–180. Also in Sovetsk. Vestn. Oftal., 1935, 6(4), 519–520.

291. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [The heightening of photic sensitivity of the eye through preliminary stimulation with light.] In S. V. Kravkov (Ed.), Iskustvennoe i estestvennoe osveshchenie v promyshlennykh predpriiatiiakh. [Artificial and natural illumination in industrial enterprises.] Moscow: Tr. i Mater. In-ta Gig. Truda i Promsanitarii NKZ RSFSR, 1935. Pp. 5–19, 71–76.

292. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [The influence of preliminary stimulation of various parts of the retina on the various sensitivities of central vision.] Dokl. Akad. Nauk SSSR, 1944, 47(7), 334–337.

293. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [On modification of differential sensitivity of central vision by means of preliminary stimulation of different parts of the retina with light.] Probl. fiziol. Optiki, 1946, 3, 67–69.

294. KRAVKOV, S. V., & SEMENOVSKAIA, E. N. [The aftereffect of illuminating the macular area of the retina with

red and green light on the photic sensitivity of rods, located at 10° and at 40° from *fovea centralis*.] *Probl. fisiol. Optiki*, 1947, 4, 65–70.

295. KRAVKOV, S. V., & SHVARTS, L. A. [On the influence of forced breathing on color vision.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 36–39.

296. KRAVKOV, S. V., & SHVARTS, L. A. [On the influence of hyperventilation on color vision.] In K. M. Bykov (Ed.), *Materialy po fiziologii retseptorov.* [Data on the physiology of the receptors.] Moscow: Medgiz, 1948. Pp. 58–62.

297. KRAVKOV, S. V., & VISHNEVSKII, N. A. [On a method of investigating night vision.] *Vrach. zan. Delo*, 1929, No. 2, 82–86.

298. KRAVKOV, S. V., & ZARETSKAIA, R. B. [Conditions of light and dark adaptation as a factor determining the direction of the eye's reaction to inadequate stimuli.] *Probl. fisiol. Optiki*, 1949, 7, 47–51.

299. KROL', TS. I. [Some problems on the photic sensitivity of the eye.] *Probl. fisiol. Optiki*, 1944, 2, 90–96.

300. KROL'-LIVSHITS. D. E. [Change in magnitude of gustatory thresholds in dogs under the influence of high external temperature.] *Arkh. biol. Nauk*, 1933, 33(2), 503–510.

301. KROL'-LIVSHITS. D. E. [The influence of muscular activity on the magnitude of gustatory thresholds in dogs.] *Tr. Vsesoiuzn. In-ta Eksper. Medits.*, 1934, 1(3), 47–51.

302. KROL'-LIVSHITS, D. E. [The influence of the temperature of solutions on the magnitude of gustatory thresholds in dogs.] *Fisiol. Zh. SSSR*, 1935, 18, 115–122.

303. KUZNETSOV, A. I., FEDOROV, N. T., & CHILAEV, A. N. [The effect of sympathomimetic amines on the electrical sensitivity of the eye.] *Probl. fisiol. Optiki*, 1948, 6, 336–347.

303a. LAZAREV, P. P. [On the mutual influence of visual and auditory excitations.] *Le Physiologiste Russe*, 1905, 4, 1–5.

304. LAZAREV, P. P. [On the mutual influence of the visual and auditory organs.] *Izv. Ross. Akad. Nauk*, 1918, 12, Series VI, 1297–1306.

305. LAZAREV, P. P. [The basic psychophysical law and its contemporary formulation.] *Usp. fisich. Nauk*, 1921, 2(2).

306. LAZAREV, P. P. Zur Theorie des Sehens. *Naturwiss.*, 1925, 13(30), 659–660.

307. LAZAREV. P. P. [Physico-chemical theory of adaptation in peripheral vision.] *Zh. eksper. Biol. i Medits.*, 1925, 1(2), 115–128.

308. LAZAREV. P. P. [The laws of action of light on the organism of animals.] *Acta Zoologica*, 1925, 6.

309. LAZAREV. P. P. [Conditioned reflexes, chronaxie, and adaptation.] *Klinich. Meditsina*, 1936, 14(1), 1–20.

310. LAZAREV, P. P. [Contemporary theory on adaptation and its application to the clinic.] In *Tr. 1 konferentsii po fisiol. optike.* [Transactions of the first conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1936. Pp. 22–26.

311. LAZAREV. P. P. [Application of contemporary theory on adaptation in the field of obstetrics and gynecology.] *Akush. i Ginekol.*, 1937, 2(4), 3–19.

312. LAZAREV. P. P. *Issledovaniia po adaptatsii* [Research on adaptation.] Moscow: Akad. Nauk SSSR, 1947.

313. LAZAREV, P. P. [On adaptation in peripheral vision under the influence of processes in other sense organs and under the influence of mental processes.] In P. P. Lazarev, *Issledovaniia po adaptatsii.* [Research on adaptation.] Moscow: Akad. Nauk SSSR, 1947. Pp. 154–158.

314. LAZAREV, P. P. [On the influence of dermal illumination on adaptation of the eye in peripheral vision.] In P. P. Lazarev, *Issledovaniia po adaptatsii.* [Research on adaptation.] Moscow: Akad. Nauk SSSR, 1947. Pp. 199–202.

315. LAZAREV, P. P. [Pregnancy and adaptation.] In P. P. Lazarev, *Issledovaniia po adaptatsii.* [Research on adaptation.] Moscow: Akad. Nauk SSSR, 1947. Pp. 234–241.

316. LAZAREV, P. P., & BULANOVA, Z. V. L'influence du travail musculaire sur l'adaptation au cours de la vision peripherique. *Dokl. Akad. Nauk SSSR*, 1936, 6(1), 5–6.

317. LAZAREV, P. P., GAMBURTSEVA, A., ABRIKOSOV, S., & SHAPOSHNIKOV, B. [On the influence of dermal illumination in man on adaptation in peripheral vision.] *Dokl. Akad. Nauk SSSR*, 1934, 4(1–2), 56–59.

318. LAZAREV, P. P., & KUPER, L. M. [Studies on the adaptation of the eye in peripheral vision.] *Arkh. biol. Nauk*, 1935, 38(3), 707–711.

319. LAZAREV, P. P., & PAVLOVA, I. KH. [On the effect of aural stimulation on the adaptation of the eye in periph-

eral vision.] *Dokl. Akad. Nauk SSSR*, 1927, Series A, No. 18, 275–276.

320. LAZAREV, P. P., PODZOROV, N. A., KUZ'MINYKH, L. V., IAKOVLEV, I. I., & MARGOLINA, E. N. [On the influence of hemorrhages on adaptation in peripheral vision.] *Dokl. Akad. Nauk SSSR*, 1933, Series A, No. 8, 138–139

321. LAZAREV, P. P., PODZOROV, N. A., IAKOVLEV, I. I., & KUZ'MINYKH, L. V. [Studies on adaptation in peripheral vision during the various stages of pregnancy.] *Dokl. Akad. Nauk SSSR*, 1934, 1(4), 182–186.

322. LEBEDINSKII, A. V. [On the dynamics of the coordinated act in the sensory sphere.] *Fisiol. Zh. SSSR*, 1935, 19(5), 945–959.

323. LEBEDINSKII, A. V. [Role of the central nervous system in the process of adaptation of the eye.] *Priroda*, 1935, 24(9), 36–44.

324. LEBEDINSKII, A. V. [Physiological regulatory mechanisms of sensitivity level of the visual apparatus.] In *Tr. 1-i konfer. po fisiol. optike.* [Transactions of the first conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1936. Pp. 27–34.

325. LEBEDINSKII, A. V. [New studies on the interaction of the retinal center and periphery.] In *Dokl. VIÉM Konfer.*, 1937.

326. LEBEDINSKII, A. V. [On the interrelation between the retinal center and periphery.] *Arkh. biol. Nauk*, 1938, 49(1), 169–175.

327. LEBEDINSKII, A. V. [Leading features in the development of Soviet physiology of the sense organs.] *Fisiol. Zh. SSSR*, 1938, 25(5), 585–617.

328. LEBEDINSKII, A. V. [Role of the central nervous system in the process of dark adaptation of the eye.] *Probl. fisiol. Optiki*, 1948, 6, 7–13.

329. LEBEDINSKII, A. V., PRESSMAN, IA., & FADEEVA, A. A. [New data on the interaction of the retinal center and periphery.] *Probl. fisiol. Optiki*, 1948, 6, 104–112.

330. LEKAKH, A. [Effect of temperature on several functions of the visual and auditory analyzers.] *Biull. eksper. Biol. i Medits.*, 1939, 8(2), 173–177.

331. LEVSHIN, V. L. [Thirty years of Soviet optics.] *Nauka i Zhisn'*, 1947, No. 14(12), 35–39.

332. LIVANOV, M. N. [On the rhythmical stimulation and interrelations of the fields in the cerebral cortex.] *Fisiol. Zh. SSSR*, 1940, 28(2–3), 172–194.

333. LIVANOV, M. N. [Biocurrents in the visual analyzer.] *Probl. fisiol. Optiki*, 1944, 2, 106–126.

334. LIVSHITS, N. N. [Dark adaptation of the eye under the influence of a field of ultra-high-frequency on the occipital region of the skull.] In *Refer. nauchn. issl. rabot sa 1944 g. Otdel biol. nauk AN SSSR.* [Abstracts of research for the year 1944. Division of Biological Sciences of the USSR Academy of Sciences.] Moscow: Akad. Nauk SSSR, 1945. Pp. 208–209.

335. LONDON, I. D. A historical survey of psychology in the Soviet Union. *Psychol. Bull.*, 1949, 46, 241–277. Also in *Psyche*, 1950, 4, 161–189.

336. LONDON, I. D. The treatment of emotions in Soviet dialectic psychology. *J. gen. Psychol.*, 1949, 41, 89–100.

337. LONDON, I. D. Theory of emotions in Soviet dialectic psychology. In *Proceedings of Second International Symposium on Feelings and Emotions: The Mooseheart Symposium.* New York: McGraw-Hill, 1950. Pp. 84–91.

338. LONDON, I. D. Contemporary psychology in the Soviet Union. *Science*, 1951, 114, 227–233.

339. LONDON, I. D. Psychology in the USSR. *Amer. J. Psychol.*, 1951, 64, 422–428.

340. LONDON, I. D. Russian psychology and psychiatry. In R. C. Christman (Ed.), *Soviet science.* Washington: American Association for the Advancement of Science, 1952. Pp. 40–47. Also in *Bull. atom. Scient.*, 1952, 8, 70–73.

341. LONDON, I. D. The Scientific Council on Problems of the Physiological Theory of Academician I. P. Pavlov: a study in control. *Science*, 1952, 116, 23–27.

342. MAAS, H. Ueber den Einfluss akustischer Rhythmen auf optische Bewegungsgestaltungen. *Arch. ges. Psychol.*, 1938, 11, 424–464.

343. MAĬZEL', S. O. *Osnovy ucheniia o tsvetakh.* [Bases of color theory.] Moscow: Gos. Izd-vo Tekhn.-Teoret. Lit-ry, 1946.

344. MAKAROV, P. O. [Study of the sensitivity of the optical system to adequate and inadequate stimulation, varying in intensity and duration.] In *Tr. 1-i konfer. po fisiol. optike.* [Transactions of the first conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1936. Pp. 245–247.

345. MAKAROV, P. O. [On the interaction of the visual organ with the auditory, gustatory, and olfactory organs.] In

I. I. konfer. to fisiol. optike. [Transactions of the first conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1936. Pp. 247–257.

346. MAKAROV, P. O. [Fluctuations in sensitivity of nervous centers, observed in periodic stimulation of the sense organs.] *Fisiol. Zh. SSSR*, 1938, 25(5), 755. (Abstract)

347. MAKAROV, P. O. [Small pendulum for studying and registering physiological and psychophysiological processes in microtemporal intervals.] *Dokl. Akad. Nauk SSSR*, 1948, 61(5), 833–836.

348. MAKAROV, P. O. [A study of nervous signalization in the visual system of man in microtemporal intervals.] *Probl. fiziol. Optiki*, 1948, 6, 173–184.

349. MAKAROV, P. O. *Neirodinamika zritel'noi sistemy cheloveka.* [Neurodynamics of the visual system in man.] Moscow: Medgiz, 1952.

350. MARGOLIN, G. I. The effect of visual and taste stimuli upon the muscular tonus in man. *Dokl. Akad. Nauk SSSR*, 1941, 33(2), 125–128.

351. MEDVEDEV, V. I. [On modification of aural, tactile, and vibratory sensitivity during dark adaptation.] *Fisiol. Zh. SSSR*, 1951, 37(1), 35–40.

352. MEDVEDEVA, N. G. [On the conditions of modifiability of color sensitivity of the eye by means of accessory stimulation.] In S. L. Rubinshtein (Ed.), *Issledovanie po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 373–383.

353. MESHKOV, V. V., & BRIULLOVA, N. V. [On a method of gauging the effect of glare on the discriminability of objects.] *Probl. fiziol. Optiki*, 1941, 1, 53–69.

354. MIKHEL'SON, M. IA. *Deistvie narkotikov na kholinesterasu.* [The action of narcotics on cholinesterase.] Leningrad: Voenno-Morskoĭ Med. Akad., 1948.

355. MILLER, I. N. [The influence of irradiation of the body on the electrical sensitivity of the eye.] *Vestn. Oftal.*, 1938, 12(5–6), 538–542.

356. MKRTYCHEVA, L. I., & SAMSONOVA, V. G. [The effect of lack of carbon dioxide and lack of oxygen on modification of thresholds of color saturation.] *Dokl. Akad. Nauk SSSR*, 1944, 44(1), 45–48.

357. MUZYLEV, F. I. [On the inhibitory influences on the retinal periphery which emanate from the retinal center.]
Fisiol. Zh. SSSR, 1948, 23(5), 753. (Abstract)

358. MUZYLEV, F. I. [On the functional dependence of the retinal processes on the retinal periphery.] *Fisiol. Zh.*, 1948, 12(?), 263–268.

359. MUZYLEV, F. I. [Further experiments on the mutual influences of the retinal center and periphery.] *Fisiol. Zh. SSSR*, 1949, 28(6), 968–841.

360. MUZYLEV, F. I. [Visual acuity as a function of intensity of illumination and level of adaptation of the eye in the discrimination of light absorption: a dark body, gray, and white objects on a light background.] *Probl. fiziol. Optiki*, 1946, 3, 59–64.

361. NALBOV, D. I. [The problem of smell and taste, the connection between them and the chemical structure of matter.] *Priroda*, 1940, 29(4), 29–37.

362. NARIKASHVILI, S. P. [On sympathetic innervation of the skeletal musculature.] *Fisiol. Zh. SSSR*, 1933, 16(3), 480–483.

363. NARIKASHVILI, S. P. [The influence of acoustic stimuli on the course of the Purkinjian afterimage.] *Izv. Akad. Nauk SSSR, Seriia Biol.*, 1944, 3, 139–155.

364. NARIKASHVILI, S. P. [The visual afterimage of Purkinje and its modifications under the influence of indirect stimulation.] Unpublished dissertation, Pavlov Physiological Institute, 1946.

365. NEISHTADT, IA. È. *Novye istochniki sveta i ikh deistvie na cheloveka.* [New sources of light and their action on man.] Moscow: Medgiz, 1952.

366. NEMTSOVA, O. [The influence of the central nervous system on several physiological processes during activity. Report No. III: On modification of the auditory threshold.] *Biull. eksper. Biol. i Medits.*, 1936, 1(6), 442–443.

367. NIKITSKII, I. N. [Modification of excitability of muscles under the influence of lights of various spectral composition.] *Biull. eksper. Biol. i Medits.*, 1938, 6(5), 611–613.

368. ORBELI, L. A. [Sympathetic innervation of the skeletal muscles, spinal cord, and the peripheral receptors.] *Vrach. Gaz.*, 1927, No. 3, 163–165.

369. ORBELI, L. A. [Survey of the theory on sympathetic innervation of the skeletal muscles, sense organs, and central nervous system.] *Fisiol. Zh. SSSR*, 1932, 15(1), 1–22.

370. ORBELI, L. A. [Basic problems in the

physiology of animals and man in the second five year plan.] *Fisiol. Zh. SSSR*, 1933, 16(2), 255–272.

371. ORBELI, L. A. [On the interrelations of the afferent systems.] *Fisiol. Zh. SSSR*, 1934, 17(6), 1105–1113. Also in L. A. Orbeli, *Voprosy vysshei nervnoi deiatel'nosti.* Moscow: Akad. Nauk SSSR, 1949. Pp. 39–53.

372. ORBELI, L. A. [On the functions of the cerebellum.] *Fisiol. Zh. SSSR*, 1935, 19(1), 255–260.

373. ORBELI, L. A. [Higher nervous activity and the adaptive-trophic role of the sympathetic nervous system and of the cerebellum.] *Fisiol. Zh. SSSR*, 1949, 35(5), 594–595.

374. ORBELI, L. A. *Voprosy vysshei nervnoi deiatel'nosti.* [Problems of higher nervous activity.] Moscow: Akad. Nauk SSSR, 1949.

375. PANKRATOV, M. A. [On the interrelationship of pain and tactile sensitivity.] *Fisiol. Zh. SSSR*, 1934, 17(6), 1238–1247.

376. POLIKANINA, R. I. [On the influence of ions of magnesium and sodium on the color sensitivity of the eye.] In *3-i konfer. po fisiol. optike.* [Third conference on physiological optics.] Moscow: Akad. Nauk SSSR, 1949.

377. POLIKANINA, R. I. [On the mechanism of action of sodium-iontophoresis on cone-sensitivity of the eye.] *Probl. fisiol. Optiki*, 1950, 9, 83–89.

378. PROPPER-GRASHCHENKOV, N. I. [Research on the eye within the Division of Physiology and Pathology of the Sense Organs of the All-Union Institute of Experimental Medicine.] *Sovetsk. Vestn. Oftal.*, 1935, 7(6), 878–882.

379. PROPPER-GRASHCHENKOV, N. I. [Interaction in the sense organs.] *Pod. Znam. Marksizma*, 1938, 17(5), 102–110.

380. PROPPER-GRASHCHENKOV, N. I. [Significance, for the physiology of higher nervous activity, of the physiology of the receptors and sense organs.] *Arkh. biol. Nauk*, 1941, 63(1–2), 3–10.

381. PSHONIK, A. G. [Analysis of cutaneous thermal reception.] *Fisiol. Zh. SSSR*, 1939, 26(1), 30–60.

382. RABKIN, E. B., & MILLER, I. N. [Problem of the technical reconstruction of Soviet ophthalmology.] *Sovetsk. Vestn. Oftal.*, 1935, 6(2), 242–244.

383. RAPPEPORT, E. IA., & ROBINZON, I. A. [On the influence of the autonomic nervous system on the visual receptors.] In *Probl. fisiol. i patol. organov chuvstv.* [Problems in the physiology

and pathology of the sense organs.] Moscow: VIEM, 1936.

384. ROSLAVTSEV, A. V. [Modifications of the angioscotomata and blind spot under the influence of thermal stimuli.] *Biull. eksper. Biol. i Medits.*, 1947, 24(4), 279–281.

385. ROSLAVTSEV, A. V. [On the influence of colored illumination on the magnitude of angioscotomata.] *Biull. eksper. Biol. i Medits.*, 1947, 24(10). No pages available.

386. ROSLAVTSEV, A. V. [Changes of angioscotomata and of the blind spot under the influence of colored illumination.] *Probl. fisiol. Optiki*, 1949, 7, 3–9.

387. ROSLAVTSEV, A. V. [On the influence of direct current on angioscotomata and the blind spot.] *Probl. fisiol. Optiki*, 1949, 7, 87–93.

388. ROSLAVTSEV, A. V. [On the altered reactions of angioscotomata and of the blind spot to chromatic stimuli.] In *Vopr. fisiol. i patol. sreniia.* [Problems in the physiology and pathology of vision.] Moscow: Medgiz, 1950.

389. RUBINSHTEIN, S. L. (Ed.) *Issledovaniia po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948.

390. RUBINSHTEIN, S. L. [Problems in the psychology of perception.] In S. L. Rubinshtein (Ed.), *Issledovaniia po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 3–20.

391. RUDNIK, A. I. [On intracentral interactions in the optic system.] *Probl. fisiol. Optiki*, 1948, 6, 265–272.

392. RZHEVKIN, S. N. [Achievements of Soviet acoustics.] *Usp. fizich. Nauk*, 1948, 34(1), 1–13.

393. SAKHAROVA, O. S. [On modification of rhythmic tactile sensations upon application of painful stimulation.] *Biull. eksper. Biol. i Medits.*, 1941, 11(3), 226–227.

394. SAMOILOV, A. IA. [Progress of ophthalmological science in the USSR after twenty years.] *Vestn. Oftal.*, 1939, 14(1), 130. (Abstract)

395. SAMOILOV, A. IA. *Iz otechestvennoi oftalmologii.* [From the history of native ophthalmology.] Moscow: Akad. Med. Nauk, 1949.

396. SELETSKAIA, L. I. [Practice and the transfer of practice in the function of brightness-discrimination.] *Biull. VIEM*, 1936, No. 6.

397. SELETSKAIA, L. I. [Modification of color sensitivity of the lateral areas of

the retina under the influence of olfactory stimulation.] *Vestn. Oftal.*, 1946, 25(6), 45. (Abstract)

SELETSKAIA, L. I. [On the effect of olfactory stimulation on visual functions at different psychophysiological levels.] *Probl. fiziol. Optiki*, 1948, 6, 319–325.

399. SELETSKAIA, L. I. [Modifiability of visual functions at different psychophysiological levels under the influence of vegetotropic stimuli.] In S. L. Rubinshtein (Ed.), *Issledovaniia po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 385–422.

399a. SEMENOVSKAIA, E. N. Weitere Untersuchungen über die Steigerung der Lichtempfindlichkeit des Dämmerungssehens durch vorhergehende Lichtreize. *Arch. Ophthal.*, 1934, 133(1), 115–120.

400. SEMENOVSKAIA, E. N. [Raising photic sensitivity of twilight vision through preliminary photic stimulation by means of red light.] *Vestn. Oftal.*, 1935, 6(4), 583–584. (Abstract). Also in *Tr. 1-i konfer. po fiziol. optike*, 1936, 231–232.

401. SEMENOVSKAIA, E. N. [The heightening of photic sensitivity after preliminary stimulation of the eyes with red light.] In S. V. Kravkov & B. M. Teplov (Eds.), *Zritel'nye oshchushchenniia i vospriiatiia.* [Visual sensation and perception.] Moscow: Sotsékgiz, 1935. Pp. 142–152.

402. SEMENOVSKAIA, E. N. [The influence of various conditions of adaptation of one eye on photic sensitivity of peripheral vision of the other.] *Vestn. Oftal.*, 1937, 10(6), 868–880.

403. SEMENOVSKAIA, E. N. [On the interaction of the retinal rods and cones.] *Vestn. Oftal.*, 1937, 11(3), 397–400.

404. SEMENOVSKAIA, E. N. [Modification of photic sensitivity of central and peripheral vision during acoustic stimulation.] *Probl. fiziol. Optiki*, 1946, 3, 94–96.

405. SEMENOVSKAIA, E. N. [On the influence of aural stimulation on subsequent photic sensitivity of peripheral vision.] *Probl. fiziol. Optiki*, 1946, 3, 97–101.

406. SEMENOVSKAIA, E. N. [Physiological stimulants and visual stereoperceptions under conditions of low illumination.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 30–33.

407. SEMENOVSKAIA, E. N. [The role of attention in modification of sensitivity of

the same organ.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 33–35.

408. SEMENOVSKAIA, L. N. [The role of attention in the modification of sensitivity of the visual organ.] *Probl. fiziol. Optiki*, 1947, 4, 148–168.

409. SEMENOVSKAIA, L. N. [Modifications in the visual analyzer on voluntary exercise of attention.] *Probl. fiziol. Optiki*, 1948, 6, 279–301.

410. SEMENOVSKAIA, E. N. [Influence of electrotonus on electrical sensitivity of the eye.] *Probl. fiziol. Optiki*, 1948, 6, 305–308.

411. SEMENOVSKAIA, E. N. [Modification of the functional mobility (lability) of the visual analyzer through influence of electrotonus under conditions of light and dark adaptation of the eye.] *Dokl. Akad. Nauk SSSR*, 1949, 68(1), 197–200.

412. SEMENOVSKAIA, E. N. [Sensitivity of the eye to direct current under influence of color-adaptation in the chromats and achromats.] In *Voprosy fiziologii i patologii zreniia.* [Problems in the physiology and pathology of vision.] Moscow: Medgiz, 1950.

413. SEMENOVSKAIA, E. N., & DUBINSKAIA, A. A. [On contrast-sensitivity of the eye in low brightness.] *Probl. fiziol. Optiki*, 1946, 3, 22–23.

414. SEMENOVSKAIA, E. N., & DUBINSKAIA, A. A. [Speeding up the process of dark adaptation.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 27–29

415. SEMENOVSKAIA, E. N., & KONDORSKAIA, I. L. [The functional lability and electrical sensitivity of the visual analyzer under conditions of red and green illumination.] *Probl. fiziol. Optiki*, 1950, 9, 35–43.

416. SEMENOVSKAIA, E. N., & LUR'E, R. N. [Changes in the electroencephalogram of the visual and lobal areas upon exercise of attention.] *Probl. fiziol. Optiki*, 1948, 6, 301–305.

417. SEVRUIGINA, M. A. [Conditioned-reflex heightening of acuity of vision.] *Vestn. Oftal.*, 1938, 12(2), 266–268.

417a. SEVRUIGINA, M. A. [Interaction of homochromatic and heterochromatic stimulation.] *Probl. fiziol. Optiki*, 1944, 2, 97–105.

417b. SHATENSHTEÏN, D. I., & KASSIL, G. N. [30 years of Soviet physiology.] *Biull. éksper. Biol. i Medits.*, 1947, 24(5), 321–347.

418. SHEVARËVA, V. K. [Conditioned sensory (auditory) reflexes.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 98–100.

419. SHEVARËVA, V. K. [The influence of

colored illumination on musculo-motor efficiency.] *Probl. fiziol. Optiki*, 1950, 9, 127–130.

420. SHEVARËVA, V. K. [The influence of colored illumination on musculo-motor efficiency in achromats.] *Probl. fiziol. Optiki*, 1950, 9, 131–133.

421. SHIFMAN, L. A. [On the interrelation of sense organs and forms of sensitivity.] In S. L. Rubinshteĭn (Ed.), *Issledovanie po psikhologii vospriiatiia.* [Research on the psychology of perception.] Moscow: Akad. Nauk SSSR, 1948. Pp. 43–93.

422. SHTERN, L. S. [Twenty-five years of Soviet biology and experimental medicine.] *Biull. eksper. Biol. i Medits.*, 1942, 14(5–6), 3–7.

423. SHVARTS, L. A. The phenomenon of sensibilization in the field of colour vision. *Dokl. Akad. Nauk SSSR*, 1944, 45(5), 217–220.

424. SHVARTS, L. A. [On the sensibilization of the apparatuses of color vision.] *Probl. fiziol. Optiki*, 1946, 3, 33–39.

425. SHVARTS, L. A. [Sensibilization to complementary spectral colors.] *Izv. Akad. pedag. Nauk RSFSR*, 1947, No. 8, 39–42.

426. SHVARTS, L. A. [Sensitivity of twilight and color vision in various emotional states.] *Izv. Akad. pedag. Nauk SSSR*, 1947, No. 8, 107–114.

427. SHVARTS, L. A. [On levels of interaction of the apparatuses of color vision.] *Dokl. Akad. Nauk SSSR*, 1948, 59(2), 405–407.

428. SHVARTS, L. A. [Modification of color sensation in emotional states.] *Probl. fiziol. Optiki*, 1948, 6, 314–319.

429. SHVARTS, L. A. [Influence of colored illumination on aural sensitivity in various states of man.] *Probl. fiziol. Optiki*, 1949, 7, 10–13.

430. SHVARTS, L. A. [On the nature of yellow light.] *Probl. fiziol. Optiki*, 1949, 7, 14–16.

431. SHVARTS, L. A. [Daily fluctuations in color sensitivity of the eye.] *Probl. fiziol. Optiki*, 1950, 9, 123–126.

432. SHVARTS, L. A. [On means of increasing the sensitivity of the sense organs.] *Probl. fiziol. Optiki*, 1950, 9, 191–196.

433. SIZOV, M. I. [On changes in sensitivity of visual thresholds under the influence of muscular effort.] *Arkh. biol. Nauk*, 1936, 41(1), 15–28.

434. SMIRNOV, A. A. [The state of psychology and its reconstruction on the basis of I. P. Pavlov's theory.] *Sovetsk. Pedag.*, 1952, 16(8), 61–88.

435. SNIAKIN, P. G. [On the physiological lability of the functions of the visual elements of the retina.] *Biull. eksper. Biol. i Medits.*, 1942, 14(5–6), 60–63.

436. SNIAKIN, P. G. [New data on the physiology of vision.] *Nauka i Zhizn'*, 1948, 15(3), 17–20.

437. SNIAKIN, P. G. [Physiological character of the so-called lability of the retina.] *Probl. fiziol. Optiki*, 1948, 6, 252–265.

438. SNIAKIN, P. G. *Funktsional'naia mobil'-nost' setchatki.* [Functional lability of the retina.] Moscow: Medgiz, 1948.

439. SNIAKIN, P. G. [Vision.] *Bol'sh. Sovetsk. Entsikl.*, 1952, 17, 211–213.

440. SOBOLEV, S. L. [On scientific criticism, innovation, and dogmatism.] *Pravda*, 1954, No. 183, 2.

441. SOKOLOV, M. V. [Problems in the psychophysiology of vision requiring solution for the civil air-force.] *Vestn. Oftal.*, 1935, 6(4), 586. (Abstract)

442. SOLOV'ËV, I. M., & FRENKEL', O. M. [On apparent movements and displacements of visually perceived objects upon stimulation of the vestibular apparatus.] *Arkh. biol. Nauk*, 1938, 50(1–2), 54–65.

443. STREL'TSOV, V. V. [Aviation medicine and physiology in the USSR after twenty-five years.] *Biull. eksper. Biol. i Medits.*, 1942, 14(5–6), 7–15.

444. STREL'TSOV, V. V., & DORODNITSYNA, A. A. [On the mechanisms of heightening photic sensitivity of the dark-adapted eye by means of ipse- and contralateral illumination.] *Biull. eksper. Biol. i Medits.*, 1942, 14(5–6), 54–56.

445. STREL'TSOV, V. V., & FEDOROV, V. I. [The influence of several pharmacological preparations on visual acuity under conditions of weak illumination.] *Biull. eksper. Biol. i Medits.*, 1944, 18(3), 63–65.

446. STROZHETSKAIA, E. IA. [The influence of sound on achromatic contrast.] *Biull. eksper. Biol. i Medits.*, 1939, 8(3–4), 279–281.

447. SUBBOTNIK, S. I. [The influence of cola and caffeine on photic sensitivity of of the eye under atmospheric conditions and hypoxemia.] *Biull. eksper. Biol. i Medits.*, 1945, 20(6), 49–51.

448. TEPLOV, B. M. [Thresholds for distinguishing brightness-differences, given different surroundings.] *Psikhologiia*, 1932, 5(3), 26–35.

449. TEPLOV, B. M. [On the interaction of simultaneous visual sensations.] In S. V. Kravkov & B. M. Teplov (Eds.), *Zritel'nye oshchushcheniia i*

vospriiatiia. [Visual sensation and perception.] Moscow: Sotsëkgiz, 1935. Pp. 2–86.

450. TEPLOV, B. M. [Inductive modification of the absolute and differential sensitivity of the eye.] *Vestn. Oftal.,* 1937, 11(1), 106–119. Also in *Vestn. Oftal.,* 1938, 13(5), 706. (Abstract)

451. TEPLOV, B. M. [On inductive modification of absolute photic sensitivity.] *Probl. fiziol. Optiki,* 1941, 1, 7–15.

452. TEPLOV, B. M. [Review of S. V. Kravkov's *The eye and its functions,* 1945.] *Sovetsk. Kniga,* 1946, No. 8–9, 82–84.

453. TEPLOV, B. M. [The sensations.] In K. N. Kornilov, *et al.* (Eds.), *Psikhologiia.* [Psychology.] Moscow: Uchpedgiz, 1948. Pp. 69–104.

454. TEPLOV, B. M. [Review of *Research on the psychology of perception,* S. L. Rubinshteĭn (Ed.).] *Sovetsk. Kniga,* 1949, No. 5, 103–108.

455. TIMOFEEV, N. N. [Modification of gustatory acuity under the influence of several physiological states.] *Fiziol. Zh. SSSR,* 1934, 17(5), 1053–1058.

456. TONKIKH, A. V. [Theory of the adaptive-trophic function of the sympathetic nervous system.] In L. A. Orbeli (Ed.), *Uspekhi biologicheskikh nauk v SSSR za 25 let.* [Progress of the biological sciences in the USSR after twenty-five years.] Moscow: Akad. Nauk SSSR, 1945. Pp. 34–39.

457. TRONOVA, A. I. [On the influence of accessory stimulation on modification of gustatory sensitivity.] *Tr. In-ta po Izuch. Mozga im. V. M. Bekhtereva,* 1940, 13, 175–182.

458. UFLIAND, IU. M. [Influence of stimulation of the dermal receptor on the functional state of the efferent and afferent systems.] *Fiziol. Zh. SSSR,* 1937, 23(1), 34–45.

459. UKHTOMSKIĬ, A. A. *15 let sovetskoĭ fiziologii.* [Fifteen years of Soviet physiology.] Moscow: Medgiz, 1933. Also in *Fiziol. Zh. SSSR,* 1933, 16(1), 1–92.

460. USHAKOV, A. A. [On training the olfactory apparatus.] *Voen.-San. Delo,* 1940, No. 1, 46–50.

461. VAVILOV, S. I. Sensitivity of the retina to the ultra-violet spectrum. *Dokl. Akad. Nauk SSSR,* 1938, 21(8), 373–375.

462. VERKHUTINA, A. I. [Electrical sensitivity of the eye and its modification with age, time of day, and several other factors.] *Probl. fiziol. Optiki,* 1948, 6, 352–359.

463. VERKHUTINA, A. I., & EFIMOV, V. V. [Electrical sensitivity of the eye and its modification under the influence of adaptation and some other conditions.] *Probl. fiziol. Optiki,* 1947, 23 : 67, 40.

464. VISHNIASSKII, N. A. & ISARLIN, B. A. [Influence of lowered barometric pressure upon adaptation, dark adaptation, and electrical excitability of the eye.] *Fiziol. Zh. SSSR,* 1935, 18 : , 247, 249.

465. VISHNIASSKII, N. A. & ISARLIN, B. A. [Dependence of photic sensitivity of the eye on prolonged examination of the brightness of the adjacent field.] *Sovetsk. Vestn. Oftal.,* 1935, 6 : , 320–329.

466. VOLKOV, N. N. *Vospriiatie predmeta i risunka.* [The perception of objects and drawings.] Moscow: Akad. Pedag. Nauk RSFSR, 1950.

467. VOLOKHOV, A. A., & DIONESOV, S. M. [Fourth conference on physiological problems, dedicated to problems in the physiology of the sense organs.] *Fiziol. Zh. SSSR,* 1938, 25(5), 753–764.

468. ZAGORUL'KO, L. T. [On the mechanism of interaction and interrelation of the afferent systems.] *Fiziol. Zh. SSSR,* 1947, 33(4), 433–447.

469. ZAGORUL'KO, L. T. [Course of the visual afterimage of Purkinje in the presence of interaction of afferent systems.] *Probl. fiziol. Optiki,* 1947, 6, 89–104.

470. ZAGORUL'KO, L. T. [On afterimages in the visual system.] *Usp. sovrem. Biol.,* 1948, 25(2), 231–251.

471. ZAGORUL'KO, L. T. [On the course of the visual afterimage of Hering and Purkinje during modification of the functional state of the nervous system.] *Fiziol. Zh. SSSR,* 1949, 35(1), 16–26.

472. ZAGORUL'KO, L. T. [On the monocular development and course of visual afterimages in the presence of photic stimulation of the other eye.] *Fiziol. Zh. SSSR,* 1949, 35(2), 143–153.

473. ZAGORUL'KO, L. T. [A criticism of the subjective method in the physiology of the nervous system and of the sense organs.] *Fiziol. Zh. SSSR,* 1953, 39(4), 498–508.

474. ZAGORUL'KO, L. T., LEBEDINSKIĬ, A. V., & TURTSAEV, IA. P. [On the influence of painful dermal stimulation on photic sensitivity of the dark-adapted eye.] *Fiziol. Zh. SSSR,* 1933, 16(5), 740–746.

475. ZARETSKAIA, R. B. [Influence of illumination of one eye with white light on the intraocular pressure of the other.]

THE INTERRELATIONS OF PSYCHOLOGY AND BIOGRAPHY

JOHN A. GARRATY
Michigan State College

This paper is an attempt to summarize the interrelationships of two seemingly separate disciplines that deal with the study of man. Biography, in all its diverse forms, attempts to describe human lives, and this is also one of the purposes of the science of psychology. Inevitably, these two fields have drawn upon each other in a manner beneficial to both, though, as will be seen, each could make still better use of the other than it does at present.

From the earliest times, the best of the writers of biography have striven to describe not only the overt actions and recorded facts of their subjects' lives, but also their personalities. The more perceptive writers have also realized that there is a connection between a man and his deeds, that an understanding of his personality helps explain his accomplishments, and that his accomplishments throw light upon his personality. This insight is implicit, for example, in these famous lines from Plutarch's life of Alexander the Great:

My design is not to write histories, but lives. And the most glorious exploits do not always furnish us with the clearest discoveries of virtue or vice in men; sometimes a matter of less moment, an expression or a jest, informs us better of their characters and inclinations, than the most famous sieges, the greatest armaments, or the bloodiest battles whatsoever (69, p. 801).

Thus, the best biographers, like the best novelists, have always been students of human psychology, and the development of psychological science has tended to confirm insights long ago achieved. The novelist E. M. Forster has written: "Psychology is

not new, but it has newly risen to the surface. Shakespeare was subconsciously aware of the subconscious" (35, p. 9). It is hardly necessary to belabor this point, which has been made many times. C. K. Trueblood, for example, has elaborately described the sound psychological principles developed by the French critic-biographer Sainte-Beuve (81). Or, to cite an American illustration, the following statements taken from James Russell Lowell's sketch of Thoreau show obvious psychological understanding:

He condemns a world, the hollowness of whose satisfactions he had never had the means of testing.

Those who have most loudly advertised their passion for seclusion and their intimacy with nature . . . have been mostly sentimentalists, unreal men, misanthropes on the spindle side, solacing an uneasy suspicion of themselves by professing contempt for their kind.

Thoreau had not a healthy mind . . . . His whole life was a search for the doctor (57, pp. 200, 205, 203).

But the tremendous development of psychology as an organized discipline since the late nineteenth century has naturally had an impact upon the writing of biography. By 1896 Havelock Ellis was calling biography a branch of "applied psychology" and bewailing the fact that most life writers knew nothing of the work of men like Wundt, G. Stanley Hall, Münsterberg, and Jastrow (32, pp. 96–97).[1] The publication of

[1] The average biographer, Ellis wrote, was producing "a figure that is smooth, decorous, conventional, *bien coiffé*, above all, closely cut off below the bust . . ." (32, p. 98). This was certainly an accurate picture of most (but not

Freud's *The Interpretation of Dreams* in 1900 and more especially of his *Leonardo da Vinci* a decade later opened up the possibility of psychoanalyzing historical figures, and of reinterpreting great masses of human history as well.

At first, the influence of the radical Freudian thinking spread slowly. With psychologists themselves hesitant or openly scoffing, biographers, if they were even aware of psychoanalysis, did not rush to adopt the technique. The earliest psychoanalytic biographies were written by European psychologists, followers of Freud, and were published chiefly in Freud's magazine *Imago*, and in other German professional journals (30). A notable exception to this was a study of Martin Luther by the American historian and Reformation scholar Preserved Smith, which appeared in 1913 and explained Luther's career in terms of the Oedipus complex (78). But Smith neither pursued the method further nor transferred his conclusions about Luther to his other writings about the Reformation. Few historians, therefore, were even aware of his effort, and thus not stimulated to copy it.

The development of the "new" history under the leadership of James Harvey Robinson after 1910, with its emphasis on broadening the interests and methods of history, did not neglect psychology. Robinson, addressing the American Historical Association in 1910, called for a general expansion of history into new fields. "History," he said, "must recognize that it is but one of several ways of studying mankind." Robinson called attention to Tarde's work in the field

of imitation, and warned his colleagues that "the relations of our reason to the more primitive instincts which we inherit from our animal ancestors . . . will never be understood without social psychology" (72, pp. 74, 93). But while the "new" history was readily accepted by historians, historical writers and biographers did not hasten to apply psychological tools to their work.

A great change in this situation came about after World War I. The experience of doctors dealing with thousands of cases of war neuroses which were obviously not the conventional "shell shock," did much to strengthen the Freudian theories. The appearance in English of Freud's *General Introduction to Psychoanalysis* in 1920, and the popularizations of psychoanalytic theory, spread the doctrine into nonprofessional areas (64, p. 329). Suddenly biographers began to see the possibilities—the insights into hidden motives, the escape from the limitations of "factual" biography, the sensationalism inherent in an approach that emphasized sex, with its healthy effect on sales. As early as 1919 Harry Elmer Barnes, a disciple of Robinson, had gone on record in support of the use of psychoanalysis in biography (9),[2] and by the next year the rush was on.

It would be of little value to attempt to list all the biographies published in the twenties that made conscious use of Freud and of his

of all) late Victorian biographies. Ellis' indictment is an interesting precursor of Lytton Strachey's famous assault on the type in the preface of *Eminent Victorians* written in 1918.

[2] Another early example of the influence of Freud on biography can be found in John Maynard Keynes's *Economic Consequences of the Peace* (1919). In discussing President Wilson's attitude toward the Versailles Treaty, Keynes wrote: "In the language of medical psychology, to suggest to the President that the Treaty was an abandonment of his professions was to touch on the raw a Freudian complex. It was a subject intolerable to discuss, and every subconscious instinct plotted to defeat its further exploration" (52, p. 49).

... within the psychoanalytical movement (use Jung and Adler (as?) a few of the most popular should be mentioned. Katharine Anthony's *Margaret Fuller* 7 was one of the earliest. R. V. Harlow's *Samuel Adams* 44, Harvey O'Higgin's *The American Mind in Action* (66), Joseph Wood Krutch's *Edgar Allen Poe* (53), Clement Wood's *Amy Lowell* 86, G. W. Johnson's *Randolph of Roanoke* 48, and L. P. Clark's *Abraham Lincoln* (21) are among the best known.

The appearance of such books as these touched off a debate among biographers, historians, literary critics, and others which rivaled the general contest between the advocates of psychoanalysis and the conservative forces of society. Naturally enough, the practitioners of the technique were loud in their own defense. Leon Pierce Clark, the biographer of Lincoln and himself a practicing psychiatrist, argued that the ordinary biographer "has not fully exhausted the possibilities of his subject, because of inadequate psychological training" (19) and cited "such competent [historical] authorities" as Robinson, Barnes, and Preserved Smith to prove that leading historians recognized the value of psychoanalysis (20).

The Freudian biographers tended to place great stress upon the close connection between a writer and his literary productions. They defended wholeheartedly their use of imaginative writings as source material for an understanding of the author's personality. Discussing Poe, Joseph Wood Krutch declared: "The forces which wrecked his life were those which wrote his books." The man and his works, Krutch argued, were "nearly identical" (53, pp. 18–19). Clement Wood, himself a poet, wrote in his *Amy Lowell:* "A poet's work

must be studied in the light of the new psychology, with the certainty that ... we will ... learn from the poetry a portrait of the artist."
86 p. 12.

Lewis Mumford, writing a "psychological" biography of Herman Melville 62, made heavy use of Melville's novel *Redburn* for material on the novelist's early life. Mumford has stated that a biographer who fails to take into account his subject's unconscious is guilty either of ignorance or childish cowardice. While admitting that the method may involve surmises based on scanty evidence, he has argued that "it is better to make mistakes in interpreting the inner life than to make the infinitely greater mistake of ignoring its existence and its import" (63, p. 5).

Perhaps the supreme tribute to psychoanalytical biography was paid by the Marxian historian Matthew Josephson:

> The dialectics of subconscious motive underlying the conscious process is to be as fruitful in measuring the individual life within reason as the dialectics of interest versus idea in the interpretation of collective society (50, p. 100).

But against such opinions as these can be placed a mass of contradictory statements, for many biographers have attacked the method. First may be mentioned those who seemed to have been motivated chiefly by prudery. It is unnecessary to do more than quote a few lines to show the nature of their argument:

> The worst foe of biography that has yet appeared is the disciple of Freud, who crawls like a snail over all that is comely in life and art (82, p. 772).
> Apart from all their sex nonsense and their ridiculous reconstruction of their characters' psychology, [psychoanalytical biographers] ... raise serious questions as to whether these authors quite realized who and what their heroes really were (1, pp. 26–27).

Actually, very few critics have resorted to such arguments as these. But many have based their opposition on a denial of Freudian principles, particularly his insistence upon the dominant importance of sex. George Bernard Shaw once wrote to Frank Harris, who was preparing a life of the playwright:

O Biographer, get it clear in your mind that you can learn nothing about your sitter ... from a mere record of his gallantries .... If I were to tell you every such adventure that I have enjoyed, you would be none the wiser as to my personal, nor even as to my sexual history (quoted in 15, p. 755).

More specifically, Emil Ludwig wrote in *Dr. Freud*, which is a bitter attack upon psychoanalytical biography: "When Freud inspects a figure out of history he resembles a physician who instead of examining his patient's whole body looks at a single organ, usually the genitals, and upon it founds his diagnosis" (58, p. 177).

Less prejudiced critics have claimed that whatever its value as a therapeutic tool, the use of psychoanalysis serves no purpose in biography. Historian Jacques Barzun, for example, has argued that the typical Freudian biography merely substitutes a new jargon for old expressions, without really throwing light on the subject's personality (11, pp. 277–278). Similarly the biographer Phillip Guedalla once wrote of "the familiar cake iced thinly with a little mental slang imported from Vienna" (41, p. 115),[3] and Hesketh Pearson has made essentially the same point (68, pp. 313–315).

A powerful complaint raised by the foes of psychoanalysis in life histories has been the lack of evidence upon which to base the study of a per-

sonality no longer living. Dumas Malone, editor of the *Dictionary of American Biography*, has put it in its simplest terms. "It is a pity," he wrote in his essay "Biography and History," "that no more authentic information is available about ... the crucial early years of nurture" (59, p. 143). André Maurois made the same point in his *Aspects of Biography*. "Who is keeping a record of Bertrand Russell's dreams, so that the Freudian biographers may interpret them at a later date?" he asked (60, p. 90). The anthropologist Alexander Goldenwiser, certainly sympathetic to psychoanalytic theory in itself, has written:

It is, of course, tempting to apply the insights of psychoanalysis to a speculative interpretation of historical characters .... But the procedure is hazardous in the extreme and, more important still, it stands in contradiction to the avowed professions of the psychoanalysts themselves (39, p. 416).

Finally, the most perceptive critics have tended to concentrate their attack on the manner in which psychoanalytical biographers have used the Freudian constructs of sublimation and reaction formation. Harry Elmer Barnes, to cite a single objectionable instance, was able to explain Alexander Hamilton's "authoritarian" personality as resulting from the fact that he was never subject to the control of a stern father. But in the same article Barnes accounted for Jefferson's anti-authority complex by explaining that his father died when he was only fourteen—before his "father image" had had time to mature! "It is significant," Barnes commented learnedly, "that Jefferson's antipathy to his father was so infantile and deep-seated that it was scarcely ever raised to consciousness." Most of Jefferson's references to his father were "reverential and awe-inspired," which to Barnes was "obvious" proof

---

[3] Psychological biography, Guedalla wrote on another occasion, is "one half of something we all knew before, and ... the other half of something which is not so" (42, p. 928).

that he hated him (10, pp. 28-37). It was this kind of analysis which led Howard Mumford Jones to complain: "So long as any symbol can be translated into any other symbol from the subject's past, so long the psychoanalyst may roam at will" (49, p. 114). Bernard De Voto, never a man to compromise a strongly held opinion, has even said: "Psycho-analysis has no value whatever as a method of arriving at facts in biography. No psycho-analytical biography yet written can be taken seriously—as fact" (26, p. 185). To which may be added the satire of Stephen Vincent Benét:

> And as for the poor devil's really loving his wife—well, I know that every authentic utterance of his on the subject says he did, and there are those dedications of his books, that sound fairly sincere. But honestly, he simply hated her all the time . . . . Don't you realize that his unconscious portrait of her as the degenerate scrubwoman with leprosy in chapter four of "Wedded and Wooed" represents his real feelings toward her? (12, pp. 169–170).

It is clear that some of the arguments against psychoanalytical biography have had great force. The initial enthusiasm of the twenties slackened off rapidly in the thirties, and while such studies still occasionally appear, they are relatively rare today. In a recent survey of British biography by V. S. Pritchett, only one book (a life of Emile Zola by Angus Wilson) was listed as utilizing "the Freudian approach," and American biography has followed the same trend (70). Examination of the *Psychological Abstracts* reveals a similar decline of psychoanalytical studies of historical personalities by psychologists. Indeed, Franklin Fearing pointed out as early as 1927 that few academic psychologists had ever actively espoused the method (33, p. 535).

The rejection of psychoanalytical techniques by most biographers has not meant a corresponding disavowal

of all psychology. Some have argued that psychology is merely "human nature" and that no formal study of the subject is necessary,[4] but the number who have paid tribute to the importance of the subject is far larger. As early as 1912 the writer Allen Johnson, while protesting against merely speculative studies of personality, called for greater awareness of the complexities of human character. The biographer should pay more heed to the early life of his **hero instead of "joyously transport[ing] the inchoate statesman into the national arena, where he will thenceforth shape the course of empire"** (47, p. 394). Johnson also suggested that an intensive study of great men might throw light on the problem of leadership, an idea that modern psychology is certainly investigating (47, p. 403).

Later writers have touched upon related points. An English critic, after denouncing the way psychoanalysis has been "ridiculously abused" by biographers, has praised the "weapon of psychology" which provides the modern biographer with "a far better surgical dissecting instrument than his predecessor had" (27, p. 128). "The new psychology," this critic writes, "has infused the art with new blood" (28, p. 19). An American student of modern biography has put it this way: "I do believe that somewhere between the extremes of the conventional method and the purely psychological there lies a mean," while Claude Fuess, an experienced practitioner of the art,

---

[4] Hesketh Pearson, for example, has denied all formal knowledge of psychology. "Dr. Freud . . . is to me but a name." To persons who have claimed that his biographies show psychological insights, his answer is that this is "simply due to my own common sense and intuition" (68, pp. 313, 315). (See also 24, p. 4.)

after examining the dangers of peering into the unconscious, concludes: "Yet with all these reservations, I am sure that the chief contribution made in recent years to the art of biography is the application of psychology to the interpretation of personality" (**36**, p. 351).

Most such statements, while well intentioned, are extremely vague. What these writers seem to be getting at in praising psychology as a tool of biography is difficult to tie down. But an effective summary has been offered by the historian Edward M. Hulme: Psychology has led to "the wholesome skepticism with which we have come to regard many of the alleged facts of history." It has made it possible to deal more satisfactorily with the conflicting statements of eye witnesses, with autobiographical writings and other personal documents, with human habits and motives—in short, "it helps to account for many facts, and also to trace the series of mental operations that come between the facts and the written record of their observation" (**45**, pp. 175–178).

Yet psychology can offer biography something more than this, though biographers have not availed themselves of it. In a few cases, where the evidence is plentiful and the conditions clear, psychoanalysis can certainly be used with justice. In the case of George III, who was clearly insane, and whose madness was thoroughly examined by contemporaries, the records are adequate, and a study such as Guttmacher's *America's Last King* is perfectly acceptable to the most conservative historians (**43**). On the other hand, it may be argued that where the ordinary evidence is extremely skimpy, as in the case of Leonardo, a psychoanalytic interpretation like Freud's may be quite use-

ful. When conventional evidence is missing, the biographer must guess in any case, and an intelligent guess based on psychoanalysis, provided it is clearly labeled a guess (as Freud's was), can surely be defended.

Even in the vast majority of biographies, which fall between the two extremes, no one should object to the use of Freudian techniques if they are explicitly described as speculations, and if known facts are not twisted or ignored in order to bring the subject into a preconceived pattern. Much of the criticism that has been aimed at psychoanalytical biography has arisen from the dogmatic way in which the authors of such works have stated their cases. As Bernard De Voto has said: "*Must* is the mechanism of psycho-analytical biography. . . . Biography proper is not concerned with the *must* but only with the *did*" (**26**, p. 188). If psychoanalytical biographers would confess to the speculative nature of their conclusions, their critics would be less bitter.

Aside from psychoanalysis, biography could benefit from other psychological techniques for the study of personality, particularly from the scientific analysis of personal documents. This is a field that has been badly neglected. When the Social Science Research Council conducted its inquiry into the use of personal documents in psychology, history, anthropology, and sociology, its historian, Louis Gottschalk, begged the question and presented a paper which confined itself almost wholly to an analysis of general historical method. "For the historian the method of dealing with human documents is all of *historical method*," he wrote (**40**, p. 6), ignoring the host of ideas adaptable to historical study in G. W. Allport's "The Use of Personal Documents in Psychological Science" pub-

lished in 1942, three years before his own work appeared. Allport's cogent analysis should be inspected by every biographer, both for its summary of the work of students of personality and for its balancing of the pros and cons of the usefulness and reliability of personal materials (5).

Certainly there are many ways in which biographers could apply knowledge gathered by workers in psychology. Recent advances in the study of graphology, for example, ought to be considered carefully. While it would be foolish to interpret a historical figure only in terms of his handwriting, the rich ore that this mine of expressive movement can yield should be dug out, refined, and forged into a handy biographical tool, for, as Allport and Vernon have pointed out, handwriting is really "brain writing" (6, p. 187). Indeed, the entire field of projective and expressive techniques, ignored by biographers, should be searched for methods that might be applicable to the study of persons no longer living.

Unfortunately, the most valuable tools, like the Rorschach and the TAT, are of no use in historical investigations, but there are others that might well be employed. Certain forms of semantic analysis of both the written and spoken word could be applied to the expressions of figures from the past, for example, D. P. Boder's discovery that there are great variations in the use of adjectives and verbs in different *kinds* of writing, and also within the writings of individuals over the course of time. Boder's study of Emerson's journals and the letters of William James brought out many interesting relationships, such as the fact that the adjective-verb ratio in James's letters varied significantly according to the sex of his correspondent (13, pp. 340–341). The literature in this field

has been summarized by F. H. Sanford (74, 75), who concludes: "Characterizations of style appear also to be characterizations of persons" (75, p. 198), a point which biographers could ponder with profit.

There are other tools that might well assist the biographer in digging out the complexities of his subject's personality, for the most important developments in this area have gone beyond the counting of mere word forms. John Dollard and O. H. Mowrer have developed a means of measuring tension as expressed, often subtly, in personal documents (29). By counting the words, sentences, and "thought units" that express some form of discomfort (D), and those indicative of relief or reward (R), they have arrived at a Discomfort-Relief Quotient (DRQ), which can be determined by the equation $DRQ = D/D + R$. Thus, hidden emotional states may sometimes be uncovered. That this device might be valuable to biographers seems obvious. (See also 51.)

Also of great potential usefulness to biographers is the study that Alfred L. Baldwin called "Personal Structure Analysis: A Statistical Method for Investigating the Single Personality" (8). Baldwin took a series of letters written by a single subject and classified them according to the goals, attitudes, and objects discussed. The resulting analysis was based on two principles: (a) The more frequently an item appears, the more important it is in the individual's personality structure, and (b) the mentioning together of items more often than would be likely according to chance indicates a relationship of these items in the personality of the subject (8, p. 168). Although a simple reading of the letters pointed up many such relationships, the mechanical counting and sorting of the ma-

terial highlighted a number of others which made the letter writer's personality (and behavior) more intelligible. If this technique was helpful to a trained psychologist, it would be even more useful for the average biographer.

Closely related to the Baldwin study is the method of "value-analysis" developed by Ralph K. White (84, 85). This technique is applicable to autobiographical writings; White developed it by studying Richard Wright's *Black Boy*. First he attempted to evaluate Wright's personality on the basis of an ordinary reading of the book. Then he applied value-analysis, going over it again classifying and counting each value judgment made by Wright. He uncovered more than half a dozen relationships, some of great importance, which he had missed in reading the book, but which, upon further thought, fitted well into the general picture he had formed (84).

The particular value for biographers of techniques like these is that they can be employed by nonspecialists. The proper recognition and classification of the material requires training it is true, but nothing like the work that must be put in by anyone seeking to use psychoanalysis. White claims that his method can be mastered in from ten to twenty hours of study and practice (84, p. 441), surely not an unreasonably long apprenticeship for a writer interested in getting inside his hero. Since these methods are plainly supplementary ones and futile to apply unless combined with subjective study, they ought not to offend the sensibilities of those who object to the employment of "mechanical" techniques in "artistic" problems. Some of these methods ought to be given at least a trial by biographers. Surely, as Saul K. Padover has written, to "illumi-

nate the internal dynamics of a person" one requires "imagination, psychological insight, psycho-social techniques" as well as the conventional documents (67, p. 16).

At the same time, life writers must beware of overemphasizing the "scientific" approach to personality. The biographer can learn from the psychologist, but he cannot reduce the *presentation* of personality to a science.

In 1924, when the rage for the "new" Freudian psychology was at its height, the American Political Science Association appointed a committee which, among other things, investigated the possible applications of psychology to the science of government. The conclusion arrived at might well be accepted as the present judgment of the usefulness of psychology in biography. "Probably the general sentiment of the committee toward psychology," reported chairman Charles E. Merriam, "would be expressed by the phrase, *con amore ma non troppo*" (61, p. 469).

Just as the first biographers were intuitive psychologists, so the early philosophers, considering the basic questions of human psychology, were influenced by what they knew of men who had lived before them as well as by their own contemporary observations. Of course, much of this information came from biographies. But it was the publication of Darwin's *Origin of Species* in 1859 which first led to serious study of the lives of men of the past by psychologists. Environmentalists saw in the principle of natural selection proof of the idea that man's surroundings have shaped his development, and sought to trace its influence in the biographies of former generations. Those who stressed the importance of heredity were driven in the same di-

reason, for by the nature of things it was difficult to study the genetic transmission of human characteristics in any other way (79, p. 67).

Since the records of ordinary men have seldom been preserved by history, most of the work that was done dealt with the extraordinary—with the "great" men of former times—and sought in one way or another to deal with the question: What is the true nature of *eminence?* Thus, Francis Galton, in his pioneering work *Hereditary Genius*, published in 1869, wrote: "Is reputation a fair test of natural ability? It is the only one I can employ . . . " (37, p. 37).[8] Galton's investigations led him to conclude that "men who are gifted with high abilities . . . easily rise through all the obstacles caused by inferiority of social rank," but that those who have all the social advantages cannot succeed "unless they are endowed with high natural gifts" (37, p. 43). Others, of course, have come to different conclusions (2, 34, 76), but perforce have turned to the same kind of biographical evidence to "prove" their points. The methods developed by Galton and those who followed him in the study of eminent persons illustrate well the use that psychologists have made of biography.

In his studies, Galton drew upon biographical collections (Foss's *Lives of the Judges*, Lord Brougham's *Statesmen of the Reign of George III*, Middleton's *Biographia Evangelica*, and other "ordinary small biographical dictionaries" which were them-

selves compilations of much biographical writing and also . . . his own words, "very many biographies" of individuals draw from the fields of politics, science, philosophy, and the arts (37, pp. 57, 106-107, 289-324). In a similar way, William James in "Great Men and Their Environment" and "The Importance of Individuals" and Frederick A. Woods in *Mental and Moral Heredity in Royalty* and *The Influence of Monarchs* extracted evidence from biographical works, and reached conclusions that, in general, followed Galton's assertion that heredity was the dominant factor producing human eminence (46, 87, 88). Charles B. Davenport's *Naval Officers: Their Heredity and Development*, a practical search for an accurate method of selecting future naval officers on the basis of "juvenile promise," made wide use of general compilations like *Who's Who*, Burke's *Peerage*, and the *Encyclopedia Britannica*, and of individual biographies whenever they were available (25, p. 3).

Investigators who rejected the *conclusions* of Galton and his followers nonetheless worked with the same kind of sources. John Fiske's reply to James's "Great Men," called "Sociology and Hero-Worship: An Evolutionist's Reply to Dr. James" (34), and a similar rebuttal by Grant Allen (2) drew from biographical information.

These studies were all more or less haphazard in their choice of evidence but with the burgeoning of comprehensive biographical dictionaries in the late nineteenth century more scientific statistical investigations became possible. This type of research was carried on by historians and biographers as well as by psychologists (Henry Cabot Lodge's "Distribution of Ability in the U. S." [55] was perhaps the earliest work of this

---

[8] Over fifty years of research did not provide a better answer to Galton's question than his own. In 1926 Catharine C. Cox wrote in *The Early Mental Traits of Three Hundred Geniuses:* "Since the problem of the early mental development of great men cannot be approached directly, because of the lack of an objective criterion of greatness, it is approached indirectly through the study of eminent men . . . " (23, p. 31).

kind), but the psychologists developed the method most fully. James McKeen Cattell, who took an intermediate position in the debate over the importance of nature and nurture, compiled his well-known list of 1,000 eminent men by computing the amount of space devoted to the individuals described in six biographical dictionaries (18). The use to which Havelock Ellis placed the British *Dictionary of National Biography* and the "over three hundred biographies of individuals" which he read in order to supplement that massive collection in writing his *Study of British Genius* (31) was essentially similar to Cattell's, although slightly more subjective. And even a much less scientific investigator, Cesare Lombroso, whose *The Man of Genius* (56) attempted to demonstrate a close relationship between genius and insanity, based most of what he had to say on information obtained from life histories.

The study of genius which made the most comprehensive use of biography was Catharine C. Cox's *The Early Mental Traits of Three Hundred Geniuses* (23). Selecting her subjects from Cattell's list, she sought facts indicative of the IQ's of these men in their personal writings and in the best available biographies. The result was both detailed and convincing, drawn from several thousand biographical works.

Many other investigators of the nature of genius have also analyzed life histories. Evelyn Raskin, in her comparison of literary and scientific leaders (71), studied *Lippincott's Biographical Dictionary*, the *Encyclopedia Britannica* and the biographical evidence contained in numerous histories of science and of literature. Similarly, Edwin L. Clarke's *American Men of Letters: Their Nature and Nurture* (22) and Cora S. Castle's *A Statistical Study of Eminent Women* (17) deserve mention in any list of works of this type, as do also the more recent *Studies in Genius* by Walter G. Bowerman (14), Mapheus Smith's "Racial Origins of Eminent Personages" (77), and R. K. White's "The Variability of Genius" (83).

Another use to which biography has been put by psychologists has stemmed from the work of the psychoanalysts. Psychoanalytical life histories, already described as an aspect of psychology's influence on biography, also illustrate the reverse; in preparing case studies of historical characters, psychoanalysts have inevitably drawn upon standard biographies. These case studies[6] are, however, essentially unlike psychoanalytical biographies. They are more methodological demonstrations than attempts to explain the subjects considered. They are meant to illustrate psychoanalytical principles, whereas psychoanalytical biographies use the principles only to explain the personalities of the characters they describe.

In the field of social psychology, some writers have made use of biography. Hadley Cantril, in his *Psychology of Social Movements* (16), studied works on Hitler to explain Nazism, and on Father Divine to explain that individual's cult. A similar but far more elaborate example of this type of work is offered by Gustave Gilbert's *Psychology of Dictatorship* (38). Although most of Gilbert's material was drawn from his interviews with captured Nazi leaders, his chapter on Hitler depended considerably upon biographies of *Der Führer*.

Other psychologists have turned to biography in an effort to discover patterns of human development.

[6] The works of this type are so numerous that it would be fruitless to attempt even an abbreviated listing of them here.

Charlotte Bühler's work with over two hundred life histories led to her conclusion that most lives can be divided into five stages: youth, period of trials, maturity, achievement, and decline (73). Similarly, Lehman's many studies of the years of peak performance in various fields of specialization illustrate an adaptation of biography to psychological purposes (see especially 54).

Biography has also been developed as a teaching technique. R. M. Elliott's course in "Biographical Psychology" at the University of Minnesota, in which each student studies biographical works about some person and writes a paper on "the success and lapses of insight with which the development of the human being is portrayed,"[7] has long been well known. Gordon W. Allport at one time adopted William E. Leonard's revealing autobiography *The Locomotive God* as a text for his course in psychology at Harvard (3), using it as "a hatrack for an entire course of study."[8] At least one other psychologist, Henry Clay Smith, then at the University of Vermont, has followed Allport's lead.

Psychologists have been neither as free nor as extreme in commenting on biography as biographers have in commenting on psychology, but they have expressed a considerable range of opinions. In general, those who have used biographies in their research have assumed the factual information contained in them to be reasonably accurate, and have tried to follow the most "authoritative" authors and to cross-check their evidence whenever possible. But they have made numerous technical criti-

cisms. Biographers seldom realize, Havelock Ellis argued, that their discipline demands "something more than pure literary aptitudes." His study of the *Dictionary of National Biography* led to him complain of a general lack of method among life writers. Often such important factors as whether or not a subject had ever married, his precise position in his family, and even so elementary a matter as a competent physical description were not to be found (31, pp. 16–18). Cox has warned that the average biographer is a synthesizer, not an analyzer, from whose work the perceptive psychologist must strip every vestige of interpretation (23, p. 24), and Evelyn Raskin has noticed that biographies of scientists fail to place enough emphasis on the personalities of the individuals they describe (71, p. 24). Alfred M. Tozzer has complained of the way many biographers distort or ignore the simplest laws of heredity in a frantic "cytological search for famous chromosomes" that will "explain" a man's fame in terms of the characteristics of his ancestors (80, p. 422).

One of the strongest criticisms of biography was made by R. K. White after completing a study of the versatility of geniuses based on Cox's data:

Biographical data of the usual type are by their very nature relatively unreliable. It is exasperating, for one who holds before himself a high standard of scientific precision, to be eternally conscious that his rock-bottom facts are not scientifically precise. On the contrary, they are often snap judgments, made by untrained observers, often by prejudiced observers, without reference to established norms (83, p. 481).

Finally, the comment of Lewis M. Terman should be mentioned. Terman decried "the utter inability of a majority of otherwise cómpetent biographical writers to appraise and interpret ... the early mental development of their subjects." He at-

tributed this to a lack of "detailed acquaintance with the age norms of childhood performance." He illustrated his point by referring to Karl Pearson's life of Francis Galton, in which the author claimed that as a child Galton had demonstrated no significant signs of precocity, despite the fact that his book described youthful accomplishments that led Terman to believe that Galton's IQ was "not far from 200" (23, p. v).

In general, biography has not made the impact upon psychology which that relatively youthful discipline has made upon it. The enthusiasm that was generated by Galton's *Hereditary Genius* has pretty generally been dissipated by time. In other psychological areas, biography has been used frequently, but in rather isolated special cases (16, 38), or in experiments that have attracted the interests of single investigators (54).

But in addition to what biography can offer to the modern psychologist by way of source material, it can also provide him with many brilliant examples of character portrayal which demonstrate the importance of literary ability in any work seeking to explain that complex organism, man.

After completing his exhaustive study of fifty men of college age, Henry A. Murray saw this point clearly:

Any conceptual formulation of man's experience . . . must necessarily do violence to human feelings. . . . This will be so because it is the substitution of heartless, denotative referential symbols for the moving immediacy of living. By employing such a scheme a person's vital movements, once warm and passionately felt, become transformed into a cruelly commonplace formula, which dispossesses them of unique value. . . . The artist's representation of an experience, on the other hand, is a re-invocation of the original . . . equally immediate, exciting and intense (65, pp. 17–18).

Murray therefore sacrificed some of the scientific objectivity of his studies on the altar of literary style. The same point has been made by Allport: "The greatest failing of the psychologist at the present time is his inability to prove what he knows to be true" (4, p. 9). Murray and Allport were suggesting that psychologists can learn from creative literature, but biography, with its stress on both artistry *and* accuracy, can, in its finest form, be a still better teacher.

In conclusion, it seems clear that biography and psychology have always been related disciplines. Early intuitive interrelationships have been supplemented in recent times by more conscious and more scientific efforts in each field to profit from the work of the other. Psychological knowledge has aided modern biographers in understanding their subjects, and biographies have been used by psychologists interested in the study of genius, heredity, social movements, and many other topics. Yet there remains room for further development. Scientific psychological methods ought to be thoroughly investigated by biographers, while psychologists should more often turn to biography as providing a source of psychological data.

REFERENCES

1. ABBOTT, W. C. Some "new" history and historians. *Mass. hist. Soc. Proc.*, 1931, 64, 286–293.

2. ALLEN, G. The genesis of genius. *Atlant. Mon.*, 1881, 47, 371–381.

3. ALLPORT, G. W. The study of personality by the intuitive method. An experiment in teaching from *The Locomotive-God. J. abnorm. soc. Psychol.*, 1929, 24, 14–27.

4. ALLPORT, G. W. Personality: a problem for science or a problem for art? *Rev. Psihol.*, 1938, 1, 488–502.

5. ALLPORT, G. W. The use of personal documents in psychological science. *Soc. Sci. Res. Coun. Bull.*, 1942, No. 49.

6. ALLPORT, G. W., & VERNON, P. E. *Studies in expressive movement.* New York: Macmillan, 1933.

7. ANTHONY, KATHERINE. *Margaret Fuller: a psychological biography.* New York: Harcourt, Brace, 1920.

8. BALDWIN, A. L. Personal structure analysis: a statistical method for investigating the single personality. *J. abnorm. soc. Psychol.*, 1942, 37, 163–183.

9. BARNES, H. E. Psychology and history. *Amer. J. Psychol.*, 1919, 30, 337–376.

10. BARNES, H. E. Some reflections on the possible service of analytical psychology to history. *Psychol. Rev.*, 1921, 8, 22–37.

11. BARZUN, J. Truth in history: Berlioz. *Univer. Rev.*, 1939, 5, 275–280.

12. BENÉT, S. V. A defense of Mrs. Anonymous. *Bookman*, 1926, 64, 168–170.

13. BODER, D. P. The adjective-verb quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, 3, 310–343.

14. BOWERMAN, W. G. *Studies in genius.* New York: Philosophical Library, 1947.

15. BOYD, E. Sex in biography. *Harper's Mag.*, 1932, 165, 752–759.

16. CANTRIL, H. *The psychology of social movements.* New York: Wiley, 1941.

17. CASTLE, CORA S. A statistical study of eminent women. *Arch. Psychol.*, 1913, No. 27.

18. CATTELL, J. McK. A statistical study of eminent men. *Pop. Sci. Mon.*, 1903, 62, 359–377.

19. CLARK, L. P. Unconscious motives underlying the personalities of great statesmen. I. A psychologic study of Abraham Lincoln. *Psychoanal. Rev.*, 1921, 8, 1–21.

20. CLARK, L. P. Unconscious motives underlying the personalities of great statesmen and their relation to epoch-making events. III. The narcism of Alexander the great. *Psychoanal. Rev.*, 1923, 10, 56–69.

21. CLARK, L. P. *Lincoln: a psychobiography.* New York: Scribner's, 1933.

22. CLARKE, E. L. *American men of letters: their nature and nurture.* New York: Columbia Univer. Press, 1916.

23. COX, CATHARINE C. *The early mental traits of three hundred geniuses.* Vol. II of *Genetic studies of genius*, Terman,

L. M. (Ed.) Stanford: Stanford Univer. Press, 1926.

24. CROSS, W. L. *An outline of biography: from Plutarch to Strachey.* New York: Holt, 1924.

25. DAVENPORT, C. B. *Naval officers: their heredity and development.* Washington: Carnegie Institution, 1919.

26. DE VOTO, B. The skeptical biographer. *Harper's Mag.*, 1933, 166, 181–192.

27. DOBRÉE, B. Modern biography. *Nat. Rev.*, 1932, 99, 128.

28. DOBRÉE, B. Some aspects of biography since Strachey. *Britain Today*, 1948, 152, 16–20.

29. DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, 42, 3–32.

30. DOOLEY, L. Psychoanalytic studies of genius. *Amer. J. Psychol.*, 1916, 27, 363–416.

31. ELLIS, H. *A study of British genius.* London: Hurst and Blackett, 1904.

32. ELLIS, H. *Views and reviews.* London: Harmsworth, 1932.

33. FEARING, F. Psychological studies of historical personalities. *Psychol. Bull.*, 1927, 24, 521–539.

34. FISKE, J. Sociology and hero worship: an evolutionist's reply to Dr. James. *Atlant. Mon.*, 1881, 47, 75–84.

35. FORSTER, E. M. *The development of English prose between 1918 and 1939.* Glasgow: Jackson, 1945.

36. FUESS, C. M. Debunkery and biography. *Atlant. Mon.*, 1933, 151, 347–356.

37. GALTON, F. *Hereditary genius.* London: Macmillan, 1869.

38. GILBERT, G. M. *Psychology of dictatorship.* New York: Ronald, 1950.

39. GOLDENWEISER, A. Some contributions of psychoanalysis to the interpretation of social facts. In H. E. Barnes (Ed.), *Contemporary social theory.* New York: Century, 1940. Pp. 391–430.

40. GOTTSCHALK, L., KLUCKHOHN, C., & ANGELL, R. The use of personal documents in history, anthropology and sociology. *Soc. Sci. Res. Coun. Bull.*, 1945, No. 53.

41. GUEDALLA, P. The future of biography. *Week-End Rev.*, 1931, 3, 115.

42. GUEDALLA, P. The method of biography. *J. roy. Soc. Arts*, 1939, 87, 925–935.

43. GUTTMACHER, M. S. *America's last king.* New York: Scribner's 1941.

44. HARLOW, R. V. *Samuel Adams.* New York: Holt, 1923.

45. HULME, E. M. *History and its neighbors.*

46. JAMES, W. *The will to believe*. New York: Longmans, Green, 1927.

47. JOHNSON, A. Tendencies in recent American biography. *Yale Rev.*, 1912, 1 (n.s.), 390–403.

48. JOHNSON, G. W. *Randolph of Roanoke: a political fantastic*. New York: Minton, Balch, 1929.

49. JONES, H. M. Methods in contemporary biography. *English J.*, 1932, 21, 113–122.

50. JOSEPHSON, M. Historians and mythmakers. *Virginia quart. Rev.*, 1940, 16, 92–109.

51. KAUFFMAN, P. E., & RAIMY, V. C. Two methods of assessing therapeutic progress. *J. abnorm. soc. Psychol.*, 1949, 44, 379–385.

52. KEYNES, J. M. *The economic consequences of the peace*. New York: Macmillan, 1919.

53. KRUTCH, J. W. *Edgar Allen Poe*. New York: Knopf, 1926.

54. LEHMAN, H. C. *Age and achievement*. Princeton: Princeton Univer. Press, 1953.

55. LODGE, H. C. *Historical and political essays*. Boston: Houghton Mifflin, 1892.

56. LOMBROSO, C. *The man of genius*. London: Scott, 1891.

57. LOWELL, J. R. *My study windows*. Boston: Osgood, 1871.

58. LUDWIG, E. *Dr. Freud*. New York: Hellman, Williams, 1947.

59. MALONE, D. Biography and history. In J. R. Strayer (Ed.), *The interpretation of history*. Princeton: Princeton Univer. Press, 1943. Pp. 121–148.

60. MAUROIS, A. *Aspects of biography*. Cambridge: Cambridge Univer. Press, 1929.

61. MERRIAM, C. E. The significance of psychology for the study of politics. *Amer. polit. Sci. Rev.*, 1924, 18, 469–488.

62. MUMFORD, L. *Herman Melville*. New York: Literary Guild, 1929.

63. MUMFORD, L. The task of modern biography. *English J.*, 1934, 23, 1–9.

64. MURPHY, G. *Historical introduction to modern psychology*. New York: Harcourt, Brace, 1949.

65. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

66. O'HIGGINS, H., & REEDE, E. H. *The American mind in action*. New York: Harper, 1924.

67. PADOVER, S. K. Architect of policy. *Saturday Rev. Lit.*, 34, October 20, 1951, 15–16.

68. PEARSON, H. *Thinking it over*. London: Hamilton, 1938.

69. PLUTARCH. *The lives of the noble Grecians and Romans*. New York: Modern Library, n.d.

70. PRITCHETT, V. S. A literary letter from London. *NY Times Book Rev.*, January 25, 1953, 26.

71. RASKIN, EVELYN. Comparison of scientific and literary ability. *J. abnorm. soc. Psychol.*, 1936, 31, 20–35.

72. ROBINSON, J. H. *The new history*. New York: Macmillan, 1912.

73. RUBINOW, OLGA. The course of man's life—a psychological problem. *J. abnorm. soc. Psychol.*, 1933, 28, 207–215.

74. SANFORD, F. H. Speech and personality. *Psychol. Bull.*, 1942, 39, 811–845.

75. SANFORD, F. H. Speech and personality: a comparative case study. *Charact. & Pers.*, 1942, 10, 168–198.

76. SCHNEIDER, J. The cultural situation as a condition for the achievement of fame. *Amer. sociol. Rev.*, 1937, 2, 480–491.

77. SMITH, M. Racial origins of eminent personages. *J. abnorm. soc. Psychol.*, 1937, 32, 63–73.

78. SMITH, P. Luther's early development in the light of psycho analysis. *Amer. J. Psychol.*, 1913, 24, 360–377.

79. STERN, W. *General psychology: from the personalistic standpoint*. New York: Macmillan, 1938.

80. TOZZER, A. M. Biography and biology. *Amer. Anthropologist*, 1933, 35, 418–432.

81. TRUEBLOOD, C. K. Sainte-Beuve and the psychology of personality. *Charact. & Pers.*, 1939, 8, 120–143.

82. WHILBEY, C. The indiscretions of biography. *English Rev.*, 1924, 39, 769–772.

83. WHITE, R. K. The variability of genius. *J. soc. Psychol.*, 1931, 2, 460–489.

84. WHITE, R. K. *Black Boy:* a value-analysis. *J. abnorm. soc. Psychol.*, 1947, 42, 440–461.

85. WHITE, R. K. *Value analysis: the nature and use of the method*. Glen Gardner, New Jersey: Society for the Psychological Study of Social Issues, 1951.

86. WOOD, C. *Amy Lowell*. New York: Vinal, 1926.

87. WOODS, F. A. *Mental and moral heredity in royalty*. New York: Holt, 1906.

88. WOODS, F. A. *The influence of monarchs*. New York: Macmillan, 1913.

CLARIFICATION OF AN AMBIGUITY IN HULL'S
PRINCIPLES OF BEHAVIOR

J. J. CLANCY, L. THOMAS CLIFFORD, AND ALLEN D. CALVIN
Michigan State College

In *Principles of Behavior* (2), Hull attempted to set forth a rigorous hypothetical-deductive system. This tremendous step forward in psychological system building has inspired more research since its publication ten years ago than any other book in the area of experimental psychology. For this reason any incongruity or ambiguity in one of the basic postulates or corollaries calls for prompt clarification.

An example of such an ambiguity occurs in Corollary VII on page 154. This corollary reads as follows: "Equal differences in the delays of reinforcement of two competing acts lead with equal practice to a lower per cent of preferential choices of the act associated with the shorter delay when the two delays are large (i.e., when the ratio is relatively coarse) than when they are small (i.e., when the ratio is finer)."

Here Hull is obviously using the terms "coarse" and "fine" to mean that the larger the fraction the coarser the ratio, i.e., 1:2 would be coarser than 1:3. However, in a preceding statement Hull notes, "Let us take, for example, the same theoretical arrangements assumed above, with the exception that the delays are of 30 and 90 seconds respectively, which gives the organism the relatively coarse ratio of 1 to 3 instead of 1 to 2, as previously" (p. 152). Again in his summary he reiterates, "Moreover, it follows from these same principles that the point of maximum ease of discrimination of a coarse ratio, such as 1 to 3, will appear at smaller absolute values than will be the case with a fine ratio, such as 1 to 2" (p. 159). This usage of coarse and fine stands, of course, in direct contradiction to the usage of these terms in the corollary cited above.

We are thus faced with an ambiguous situation, which can be resolved in two possible ways. We can reverse the words "large" and "small" and change the corollary to read: Equal differences in the delays of reinforcement of two competing acts lead with equal practice to a lower per cent of preferential choices of the act associated with the shorter delay when the two delays are small (i.e., when the ratio is relatively coarse) than when they are large (i.e., when the ratio is finer); or we can reverse the words "coarse" and "fine" so that the corollary reads: Equal differences in the delays of reinforcement of two competing acts lead with equal practice to a lower per cent of preferential choices of the act associated with the shorter delay when the two delays are large (i.e., when the ratio is relatively fine) than when they are small (i.e., when the ratio is coarser).

These two alternatives are diametrically opposed in meaning. If we interchange the words large and small, as we did in the first alternative, the corollary states that there will be a lower percentage of choices when the two delays are small, i.e., when the ratio is relatively coarse. If we interchange the words coarse and fine, as we did in the second alternative, the corollary states that there will be a lower percentage of choices when the two delays are large, i.e., when the ratio is relatively fine.

Fortunately Hull provides us with the answer as to which is the proper alternative to choose. He states that this corollary finds empirical confirmation in a study by Anderson (1). An examination of Anderson's data, which Hull presents, indicates that the second alternative is the one that is in accord with these findings. Thus it seems apparent that this corollary of Hull's should be corrected to read: Equal differences in the delays of reinforcement of two competing acts lead with equal practice to a lower per cent of preferential choices of the act associated with the shorter delay when the two delays are large than when the ratio is relatively fine, than when they are small (i.e., when the ratio is coarser).

REFERENCES

1. ANDERSON, A. C. Time discrimination in the white rat. *J. comp. Psychol.*, 1932, 13, 27–55.

2. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

A REJOINDER ON ONE-TAILED TESTS

LYLE V. JONES
University of Chicago

In a recent issue of this journal, Burke (1) criticizes earlier discussions of one-tailed and two-tailed tests (3, 4, 5) and suggests need for caution in the application of one-tailed statistical tests to psychological research designs. The writer is in accord with several implications of Burke's note. As is true for most statistical designs, abuses would be reduced markedly were the test model completely specified and justified in terms of the purpose of investigation, before data are viewed by the investigator. To be guided by the data in the specification of hypotheses and statistical tests is a grave breach of the rules of experimental verification (cf. 6).

The argument presented by Burke, however, is more than a plea for careful consideration of the choice of test models for given experimental problems. It is stated that the selection of a one-tailed test model requires that an investigator be willing to "publicly defend the proposition . . . of no difference" if results actually show a large difference in the direction opposite to that predicted (1, p. 387). The proposition appears indefensible, since it demands arguing that a particular observed difference, no matter how large, is only a sampling departure from zero. If accepted, Burke's argument should lead to universal avoidance of one-tailed tests.

Consider the following experimental problems, selected for simplicity as single variable designs: (a) On the basis of a certain behavioral theory, we might predict that an experimental condition imposed upon subjects in the population under study would raise the mean level of performance on a given task. The theory provides a prediction; an alternative hypothesis that the mean for the experimental group population will exceed the mean for the control group population. The hypothesis under test is that the mean for the experimental population is the same as or less than that for the control population. We should like a statistical test that will yield a decision: either the data are consistent with the hypothesis under test, or we reject the hypothesis in favor of the alternative. (b) In a field of applied psychology a new diagnostic technique is developed and is to be adopted if, and only if, we are confident that it is better than the current technique which would be replaced. Assuming the availability of a suitable criterion, the two techniques are applied to comparable samples or to the same sample; interest resides in the extent to which the parametric proportion of successful predictions using the new technique exceeds that using the old. A statistical test is to supply a decision: either the new technique is no more adequate than the old, or the new technique is more adequate.

For the class of problems illustrated by these two examples, the hypothesis under test is not simply one of no difference. We wish to test the hypothesis that the algebraic difference between parametric mean performance under experimental and control conditions is zero *or negative* against the alternative hypothesis

that the difference is positive. It is meant to stress this formulation of the one-tailed statistical test with the greatest possible emphasis. With this formulation, it is apparent that acceptance of the hypothesis under test does not demand defense of a proposition of no difference; the observed difference, whether negative, zero, or slightly positive, simply does not allow acceptance of the alternative hypothesis at the level of stringency (values of α and β) chosen for the test.

In a footnote, Burke (1, p. 385) criticizes this formulation as it appeared earlier (4, p. 45) on the grounds that it confuses hypotheses to be tested with hypotheses to be guarded against as alternatives. To the contrary, this statement of the problem clarifies the nature of the hypothesis under test. The hypothesis to be tested is[1]

$$H_0: \mu_e - \mu_c \leqq 0,$$

where μ_e is the population mean for the experimental condition, μ_c is the population mean for the control condition, and the experimental prediction yields an alternative hypothesis,

$$H_1: \mu_e - \mu_c > 0.$$

Burke's primary argument seems to rest upon the contention that "there is, under the single-tailed rule, no safeguard whatsoever against oc-

[1] An equivalent formulation of the one-tailed test model, of which the writer was unaware at the time of his earlier note, is that proposed by Dixon and Massey (2, pp. 100–104).

casional large and serious errors when the difference is in the unexpected direction" (1, p. 385). Our formulation of the hypothesis under test completely resolves this difficulty, for no error is committed when that hypothesis is accepted on the basis of a large observed difference in the unexpected direction. The event is one of a class of events consistent with the hypothesis tested.

If one were to retain the alternative hypothesis, H_1, above, and to adopt a two-tailed statistical test, accepting H_1 when there were observed large differences between the means *in either direction*, his position would be unenviable. For, following the rules of his test, he would have to reject the hypothesis under test in favor of the alternative, even though an observed difference, $\mu_e - \mu_c$, was a substantial negative value.

The remaining discussion by Burke consists of pragmatic arguments against the adoption of one-tailed tests. The arguments appear valid only under the assumption that every application of the one-tailed test is an abuse of experimental methodology. Certainly, if (a) the test model is specified completely (including specification of the confidence level to be adopted) before the data are gathered, and if (b) the purpose of the test is only to determine whether a particular directional prediction is supported by the data, then the one-tailed test not only is appropriate, but it is in error to use a two-tailed test model.

REFERENCES

1. Burke, C. J. A brief note on one-tailed tests. *Psychol. Bull.*, 1953, 50, 384–387.
2. Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
3. Hick, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, 59, 316–318.
4. Jones, L. V. Tests of hypotheses: one-

sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, 49, 43–46.
5. Marks, M. R. Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Rev.*, 1951, 58, 179–184.
6. Marks, M. R. One- and two-tailed tests. *Psychol. Rev.*, 1953, 60, 207–208.

FURTHER REMARKS ON ONE-TAILED TESTS[1]

C. J. BURKE
Indiana University

In the discussion of one-tailed tests there are a number of issues which have not been clearly separated. An effort will here be made to separate them.

Two Types of Models

Psychologists have used two different types of models in interpreting their data: statistical models and specifically psychological models. The two types have been employed in different ways and for entirely different purposes. Statistical models have been used to determine whether data are intrinsically interesting or, in other words, whether the data indicate relationships between psychological variables. Psychological models have been utilized in attempts to organize the knowledge attained in various areas of the field.

Early uses of statistical methods mixed the psychological and statistical models. The substantive biases of the experimenter led him to seek out, with each set of data, those statistical techniques which supported his views and to neglect the techniques which did not. Gradually, it was recognized that the function of the statistical model was prior to and separate from the function of the substantive model. Cognizant of the fallibility of their data and the greater fallibility of their theories, psychologists developed the view that data should be interpreted without the intrusion of the biases, however well founded, of the experimenter

and accordingly accepted certain statistical procedures as conventions. The advocates of one-tailed tests would take us a large step backward, for they openly favor the mixing of the statistical and substantive models.

Any psychological theory worth bothering with will generate a number of predictions which can be directly checked against experimental data. Such direct determination of the consistency of data with the model is the only proper test of a psychological theory. In a check of this kind, there is a point by point verification of the theory, and strictly speaking, no statistical model is necessary. This does not mean that there should be no statistical interpretation of the data; it is usually wise to make the ordinary statistical interpretation prior to the theoretical check—to determine whether the data show enough to warrant a theoretical analysis. If, in any given experiment, a theory provides only a directional prediction, there seems to be nothing wrong with our traditional procedure of first establishing the presence of a difference between groups and subsequently noting its direction. A statistical model should enter directly into the verification of a substantive theory only when it is based on a null hypothesis with respect to the residuals of the data from theoretical predictions.

A statistical model can be viewed as a technique for assessing the reliability of an experiment without repeating it. The question answered by a statistical test is whether or

not the sample obtained is a member of a certain class of samples. This class is usually defined by the null hypothesis to be tested and by the conditions under which the experiment would be repeated, if repetition were undertaken; frequently certain supplementary parametric assumptions must be made. Since an experiment can be repeated without any appeal to a body of substantive theory, no such appeal need ever be made for purposes of statistical analysis. Experiments are often designed from theoretical considerations, and the conditions of repetition may, therefore, be dictated by a substantive model. In some analysis of variance designs there may be freedom in the choice of an error term. This freedom indicates that the single experiment which has been performed, even in conjunction with a null hypothesis, does not define a unique class of repetitions. When this is the case, appeal to substantive theory may lead to the choice of a proper error term, and this appeal may seem to involve a direct intrusion of substantive theory upon the statistical interpretation. As we have put the matter, however, it is readily seen that in this instance the theory does no more than define the conditions under which the experiment would be repeated.

The One-Tailed Test in Theoretical Statistics

Properly described, the concept of a one-tailed test is clear and free from any objection on mathematical grounds. There is no disagreement on this point. Disagreement on what constitutes a proper description of a one-tailed test centers specifically on the nature of the hypothesis that is tested. Jones has emphasized his conviction that the hypothesis of a nonpositive difference is tested against the alternative of a positive difference. I have stated that a null hypothesis is tested with safeguards against only a subset of the possible alternatives.

As it now stands, the mathematical theory of testing statistical hypotheses requires that the hypothesis under test lead to the calculation of a level of confidence. (Until the onslaught of the one-tailers, books in psychological statistics cautioned us to test only "exact" hypotheses.) If we ask Jones to exhibit his calculation of the level of confidence in a one-tailed test, we shall probably discover that he bases the calculation on the distribution obtained under the null hypothesis. Then, according to the usual statistical logic, he is in fact testing the null hypothesis against selected alternatives. The comments of my earlier note were made within this framework.

An easy modification of statistical logic can make the one-tailed test of a nonpositive difference tenable. Since the probability that he rejects the hypothesis of a nonpositive difference when it is true can, in his procedure, never exceed the level of confidence calculated under the null hypothesis, Jones can declare that he has a bound on his level of confidence even though he does not know its precise value. Utilizing this bound, he can set up consistent rules for one-tailed interpretation of the hypothesis he has stated.

The two paragraphs above present what seem to be the only two positions open to Jones. Whichever he accepts, I am in complete accord with him so far as mathematical statistics goes, for both are clear and statistically defensible. Yet his succinct assertion that I advocate universal

avoidance of one-tailed tests is correct. The locus of disagreement is to be found in practice, not in theory.

APPLIED PROBLEMS

In applied problems like Jones's example of the drugs, it is often true that certain practices or routines will be changed if, and only if, an experiment yields a reliable difference *in a specified direction* if the new is reliably better than the old. For problems like this, Jones maintains that the two-tailed test is inappropriate. Verbally his argument proceeds well, but before we accept it, let us note that problems of this kind have never given any difficulty—we have always understood what we did and why. Let Jones use his one-tailed test at the 5 per cent confidence level to analyze data from a large number of such experiments. We shall analyze the same data using a two-tailed test of the null hypothesis but permitting ourselves a luxuriously loose 10 per cent confidence level. Being sensible men, we shall not advocate a change to the new routine if the old one proves reliably superior. Unless we wish to differentiate a null from a negative difference (we might, for example, consider a replication necessary unless the null hypothesis were rejected), the two procedures will obviously lead to the same recommendations with every set of data. Thus, they have identical consequences for the immediate practical decision.

For myself, I do not much care whether one uses a one-tailed or two-tailed model in such applications. Since the choice of a level of confidence is arbitrary and since identical decisions will always be made, we have complete equivalence between the one-tailed model and a doubled level of confidence in our traditional procedure. It would be wrong for either Jones or me to protest that compelling reasons exist for preferring one of those ways of talking about applied experiments over the other. Any appeal to the risks of the two kinds of errors is little more than an uninteresting parlor game. Since no gain results, there seems to be insufficient reason for changing the traditional two-tailed interpretation, but this is not a very strong objection. In my earlier note, I conceded the admissibility of the one-tailed test for the experimenter's own laboratory planning; I am quite willing to add the applied experiment to the list.

THE PUBLIC LITERATURE

There do seem to be compelling reasons why the use of one-tailed tests in our permanent reports must be rejected. These reasons are of two kinds: one kind residing in the nature of the scientific enterprise and the functions of the experimental literature, and the other associated with the debasing effect upon the literature that is almost certain to result from adopting one-tailed tests.

Experimental scientists must have for data a permanent respect that transcends their passing interest in the stories they make up about their data. Science is a public enterprise. These are statements to which most experimental scientists would subscribe. One-tailed tests violate both the spirit and the letter of these statements. Interest in whether a relationship is found between two variables in a given experiment is seldom confined to those who share the theoretical preconceptions of the experimenter, and the right to discuss experimental data in relation to a particular theory does not remove the obligation to interpret the experiment according to rules

that can be accepted by the reader who rejects your views as well as by the reader who shares them. Our experience has shown that experimental results can be viable for years after they are first reported even when the experiments are designed on the basis of moribund theoretical considerations. Let us consider our obligation to future psychologists who might not be able to understand why we were so naive in 1954 as to predict *that* direction for our experimental results. Why should we put needless difficulties in the way of any reader, present or future?

Why, particularly, when the change over to one-tailed test is almost certain to gain nothing and lose much? Jones asserts that my dire forecasts on the effects of one-tailed tests on our literature are predicated on the assumption that every one-tailed test is a misuse. Here he has missed the fundamental point of my argument. As in applied experiments, there is pragmatic equivalence between using the one-tailed test and doubling the level of confidence in the two-tailed test, for if the discrepancy is in the expected direction, the power of these two tests is identical, and experiments should be, and usually are, set up with power as a foremost consideration.

Most psychologists would resist the general adoption of the 10 per cent level for rejecting hypotheses. Such a low level of aspiration would

lead to a decreased number of subjects per experiment and thus accomplish a gradual attrition in the soundness of our reported conclusions. This is precisely what will happen if there is general, and editorial, acceptance of the one-tailed test.

CONCLUDING REMARKS

A diminution in the quality of our psychological literature may be of little moment to the theoretical statistician but must be of grave concern to the experimental psychologist. Once they have the assurance of the mathematical statistician that the procedure is logically defensible, experimenters seem so eager to employ one-tailed tests that, sociologically speaking, I have lost the argument almost before it has been joined. Of course, the procedure is mathematically sound, but psychologists are mistaken if they believe the mathematical statistician can speak with special authority in the matter. The decision is for us alone; it is for us to ponder whether, in the interpretation of our data, the one-tailed test is wise. We are drifting into an unfortunate decision, and I, for one, wish to enter a plea for a little less methodology and a little more wisdom. We expound a distinction between experimental hypotheses and statistical hypotheses to our beginning students; let us maintain the distinction for ourselves.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology* (Rev. Ed.) New York: Holt, 1954. Pp. xi + 948. $8.95.

It is a pleasure indeed to note the appearance of this revised edition of one of psychology's classics, Woodworth's *Experimental Psychology*, which has graced the shelves of every graduate student for a generation. In a real sense the tradition is even longer, for the spirit of this revision is very much the spirit of Ladd and Woodworth's *Physiological Psychology* which an older generation will remember (and perhaps that of the still older Ladd, which this writer willingly confesses he does not remember!). The revised edition is a concretely objective account of the principal facts discovered by psychologists in the areas of sensation and perception, learning, and the emotions, with frequent appeal to established physiological mechanisms for explanation.

It should be emphasized at once that this is truly a revision and not just a freshened-up second edition. While the range of material, and even many of the chapter headings correspond closely to the Woodworth book, the treatment of some topics has undergone very substantial change. Learning, in particular, has been almost completely reorganized. Three of the old chapters have disappeared: Retention, Memory for Form, and Practice and Skill. Three new chapters have been added: Introductory Survey, Discrimination Learning, and Motivation in Learning and Performance. New names appear: Hull, Spence, Skinner, Tolman, Hilgard, Hovland, while older names take second place, G. E. Müller, Ebbinghaus, Pavlov, and Carr.

Changes in other sections of the book are less drastic. Feeling, Experimental Esthetics, and Reading are not given separate treatment, and the chapters on Problem Solving and Thinking are combined into one shorter chapter. The subject of emotion is given a new, substantially more physiological treatment. Of the sensory chapters, Vision, in particular, reflects the renewed emphasis which the war gave to this side of experimental psychology.

It is always tempting to a reviewer to seek some measure of the amount of change which has been made in revising a book. In this case, a list was made of the most frequently cited (indexed) authors. Highest ten in the 1938 book were, in order, G. E. Müller, Wundt, Ebbinghaus, Köhler, Cattell, Dodge, Judd, Thorndike, Pavlov, and E. H. Weber. Of these, only Thorndike survives in the 1954 list. The list now runs Hull, Boring, Hilgard, Thorndike, Stevens, Lashley, Graham, Wever, Hovland, Tolman. Quite a change!

The present book is roughly 25 per cent larger, yet less attention is given to any one person. This makes doubly surprising the fact that Hull is cited almost twice as frequently as any single person in the 1938 edition. If anything is suggested by this tabulation, it is that there is considerably more dependence in this book on secondary sources. The frequent references to Hull, Boring, and Hilgard, for instance, are rarely to original papers, but much more to their organized judgments about the facts in a given experimental field.

The critical review of a book of this kind is not easy. It contains too many detailed facts. The perspectives are rarely global. There is the added uncertainty of which of the two authors has seen a problem in a given way. It would be easy, therefore, to add some comment at this stage recommending this book for its undoubted excellence, and advising teachers and librarians that it merits their special attention. Both comments are true. But the real test of these propositions lies in the popularity which *Experimental Psychology* has enjoyed over the past 16 years, and in whatever degree of acceptance the revised edition may find in the next decade. It is of much more interest to see the kind of reflection which the book affords us of the contemporary state of affairs in experimental psychology.

1. The strongest impression of this reviewer is the striking unevenness in the development of various parts of the field. Some of this may proceed from the very catholicity of interest on the part of the authors. Or it may be that there are larger gaps between older and younger generations of psychologists than some of us would care to admit. Nevertheless, the diversity in approach, in sophistication as to method, and in the formalization of results strikes the reader repeatedly in trying to read such a book as this from front to back. The first three chapters (skipping an Introduction) on Reaction Time, Association, and Attention present almost sheer anachronism. The problems, objectively conceived, certainly are problems with which the experimental psychologist wants to deal, but scarcely in this intellectual framework. But the authors' trouble was obviously how to recast the older facts to fit them to a currently ac- ceptable pattern. Where, for instance, is the body of modern evidence to which old facts about set and attitude and the selectivity of attention should be related?

To take another contrast, there is the curious difference between the field of learning on one hand and the sensory fields on the other. There is no question that great changes have taken place in the field of learning. To this the revised edition is ample testimony. Neither "reinforcement" nor the "law of effect" can be found in the 1938 index; their omission in the 1954 book would be unthinkable. And yet the field of learning is full of doctrinal controversy. Is there one, or are there two kinds of conditioning? Can we depend on the shape of the generalization gradient? What about the conflicting views of the nature of extinction? In short, theory building and the effort at semantic clarification seem to be the only analytic tools at hand. In contrast, nothing frightens the sensory psychologist so much as a bit of theory. His experiments are brutally detailed; their plan, entirely parametric. For explanation, he takes flight into physiology. Is this, rather than theory building, the syndrome of a mature science? Certainly, as presented in the revised edition, these two areas of interest follow an utterly different pattern, and the difference is puzzling indeed.

2. Not unrelated to the unevenness in development and in method is the lack of an overview of what psychology's problems are, where they come from, and to what consequences they lead. Like Topsy, most of the questions in this book "just growed." For the expert or for the technician this deficiency is unimportant. He does not go to a factual compendium such as this to answer the question

about why, as a psychologist, he wants to know this or that. But it is not entirely unimportant for the graduate student who reads a treatise of this kind. He is more likely to be forming some notion of the way in which an experimental psychologist regards human experience and human behavior. One thing this book makes clear—the methods to be followed are experimental. Concrete methods are repeatedly detailed. But often the reader, at least this reader, gets the impression that we drift into experimental work on an issue without keeping our original objectives in mind.

Let us take the question of the GSR and emotion. Very roughly the conclusion seems to be that the GSR is probably related to something called the level of activation. Or at least the GSR reflects changes in the level of activation. But the GSR does not have much to do with what has traditionally been called emotion and that was the broad class of behavior with which the discussion started. Perhaps the term emotion is just another dead dodo, but the reader may feel that the GSR is a Trojan horse that takes the field without a battle. Too often, the discussion slips from a problem to a method, and from the method to controversy about the method. This latter is not always psychology.

3. When psychology deals with questions asked as often for methodological as for substantial reasons, systematic comparisons of the findings in related fields are likely to be lacking. For instance, a great deal of the discussion of vision hinges around stimulation with short flashes of light. The parallel experiments on the temporal course of stimulation in other sense modalities come in for rather little discussion and the possi-

bility of a common underlying mechanism does not come up. The same kind of question is, however, recognized with regard to adaptation—and remains unanswered. Perhaps the contrast is most striking if one recalls the almost overelaborate treatment of interrelated experiments in Koffka's *Principles of Gestalt Psychology*, or the parallel effort in Hull's *Principles of Behavior*. It is indeed a wise man who can steer the way between the Scylla of theoretical vacuity and the Charybdis of a mere taxonomy. This reviewer would argue that the authors have sailed a bit too close to the whirlpool.

It would not be proper to close this review without a word of appreciation for the real scholarship that lies behind this revised edition, quite as much as it did in the case of the 1938 *Experimental Psychology*. Everyone who sits down with the book, though he will miss some details of topics near to his own interests, may as often find the discussion cogent and complete. But rarely, very rarely, will some mention of an important line of research be entirely lacking. Within certain limits as to subject matter, there is amazingly little that has not been at least considered by the able authors. This in itself is no mean achievement.

EDWIN B. NEWMAN
Harvard University

WEITZENHOFFER. A. M. *Hypnotism: an objective study in suggestibility.* New York: Wiley, 1953. Pp. xvi+380. $6.00.

A good deal of experimental work in the fields of suggestibility and hypnosis has been done since Hull published *Hypnosis and Suggestibility: An Experimental Approach*. A number of books have been published in this field too, but they have usually

been of a speculative character or have dealt only with limited aspects of the field. *Hypnotism: An Objective Study in Suggestibility* is the first real attempt to pull together all the available experimental material, to assess it critically, and to draw whatever conclusions may be justified on the basis of the findings. The attempt was well worth making, and the scientific care and objectivity of the author have succeeded in making this an indispensable book for anyone interested in the theory and practice of hypnosis, or doing research in this field.

The book is divided into four parts, the first of which gives a general orientation and deals with such topics as the criteria and stages of hypnosis, tests of suggestibility, and so forth. The second part deals with experimental studies of the *intrinsic* and part 3 deals with the *extrinsic* characteristics of suggestibility and hypnosis. This differentiation is not clearly defined by the author. By intrinsic he seems to mean such things as the factors influencing suggestibility, the interrelationship of hypnosis and suggestibility, and "animal hypnosis." By extrinsic he seems to mean the effects of suggestion and hypnosis on sensory and perceptual functions, on memory and learning, on organic functions, and so forth. Some of the chapters included in one part, however, would seem to fall more easily into the other, if this be indeed a correct interpretation of the author's wish; altogether this division does not seem to be a very happy device. The last part deals with a review of theoretical formulations and presents the author's own theory.

Weitzenhoffer accepts the division of suggestibility into primary, secondary, and tertiary, and builds his account around these notions of inde-

pendent factors. In doing this, however, he is not always consistent. Thus, on page 33 he quotes with approval data showing that neurotics are more suggestible on a test of primary suggestibility; a few lines later he quotes another investigator's results showing that neurotics are more suggestible on a test of secondary suggestibility. Weitzenhoffer concludes that the latter *confirms* the earlier work. Now if primary and secondary suggestibility are indeed independent factors, a test of one type cannot itself confirm results obtained from a test of the other type. This tendency to have his cake and eat it too (i.e., to regard two factors as independent, but yet to assume that tests of one type can confirm results obtained by tests of the other) occurs several times throughout the book and may be misleading to readers not familiar with the meaning of orthogonal factors.

On most occasions, Weitzenhoffer's conclusions from the evidence are unexceptionable and will be accepted by most students of the subject. Occasionally, however, he seems to miss the mark. Thus, in his summary of the literature dealing with the relationship between neurosis and suggestibility he ascribes the contradictions found to the fact that differences are observed when psychiatric diagnoses are used, but not when neuroticism inventories are used. A different summary might be that differences are found when severe neurotics are compared with normal people, but not when we restrict the range so much that only normals are included in the sample.

The documentation of the book is very thorough indeed, the bibliography running to over five hundred titles, all well selected. The reviewer has been able to find only one major

emission that would alter the conclu-
sions Weitzenhoffer draws. In his
discussion of hypnosis and dissocia-
tion (p. 99). Weitzenhoffer quotes the
work of Mitchell which is negative
but not that of Nagge which is posi-
tive. Since it is derived from a larger
number of cases, the latter appears
more trustworthy and would have
altered the conclusion drawn if it had
been considered. This is important,
for it determines the author's atti-
tude to the dissociation hypothesis.

On much disputed points like the
existence of animal hypnosis, or the
possibility of inducing antisocial acts
under hypnosis, the author sums up
with impartiality and bases his con-
clusion entirely on the evidence with-
out showing any tendency to try to
make the evidence agree with pre-
formed conclusions. This tendency to
follow the facts rather than to try to
press them into the service of some
general theory makes this quite an
unusual book in the field.

The major new contribution of the
book lies in the theory advocated by
Weitzenhoffer to account for the
phenomenon of hypnosis. This is a
three-stage theory: "*Hypnosis is the
outcome of not one but of several tem-
porally overlapping and interacting
processes.*" The two main ones the
author considers to be those of *homo-
action* and *generalization;* in the
later stages of hypnosis a third
process, namely a *narrowing of con-
sciousness*, is introduced. "All of
these complement each other, and are
interwoven, although each appears to
have a different mechanism of action
and different end results when con-
sidered singly." This theory is too
complex to be considered in any de-
tail here. It seems to have the usual
advantages and disadvantages of
theories involving the interaction of
several processes, none of which can

be verified or measured with any de-
gree of accuracy. It would probably
be possible to account for most psy-
chotic phenomena using the method
this hypothesis. But it would be very
difficult to make predictions which
could invalidate it. Another weak-
ness of this theory seems to be that
although certain aspects of the theory
involve learning and presumably also
drive reduction, the author has no-
where tried to bring his formulation
into line with modern learning theory.
Such an attempt might have resulted
in testable predictions, thus strength-
ening the acceptability of the theory.
In spite of these criticisms, theoreti-
cal formulations in this field have
generally been so obviously untena-
ble that Weitzenhoffer's formulation
stands out as a much more reasonable
interpretation of the very complex
phenomena under discussion than
any alternative theory known to the
reviewer. Perhaps it would be unfair
to ask for more than that at this
stage of progress.

H. J. EYSENCK
University of London

ROE, ANNE. *The making of a scien-
tist.* New York: Dodd, Mead &
Co., 1953. Pp. ix + 244. $3.75.

Since scientific discoveries have
played a central role in initiating the
vast changes in human life which
have characterized the past two cen-
turies, one cannot doubt the impor-
tance of obtaining better understand-
ing of creative investigators. Dr.
Roe for several years has been con-
ducting studies of research scientists.
She has obtained test data as well as
biographical and personal interview
data from a considerable number of
the leading scientific men of today.
Her original reports of her findings
have appeared in monograph form
and in psychological journals. The

present volume is a summary account of her investigations written for the layman. It contains no new data. The style is informal and easy. However, the reviewer is inclined to believe that most readers who are interested in "the making of a scientist," will find the presentation too elementary. Furthermore, if the reader should become interested in delving further into the characteristics of scientists, this book will not assist him, for he will find in it few references to other publications, including Dr. Roe's own research reports. This is unfortunate, for many psychologists, as well as some general readers, would find a bibliography useful.

The critical reader will doubtless take issue with some of the conclusions which Dr. Roe arrives at from studies of quite small groups of subjects whose representativeness is unknown. However, all in all, Dr. Roe is to be commended for taking the pains to make available to interested laymen an account of her work. Psychologists too seldom undertake to communicate with other than psychologists.

WAYNE DENNIS

Brooklyn College

DOLLARD, JOHN, AULD, FRANK JR., & WHITE, ALICE. *Steps in psychotherapy*. New York: Macmillan, 1953. Pp. ix+222. $3.50.

For reasons which are no longer valid, psychotherapists for over half a century have failed to make available for public inspection the raw data from which their hypotheses and conclusions have been derived. The procedure of relying upon the subjective interpretations of an ego-involved therapist-investigator is gradually being replaced, fortunately, by typescripts of phonographic

recordings. Although the present book presents only one verbatim interview from a series of 17 contacts with a patient, with the remaining interviews being abstracted from recordings, the reader gains a definite impression of intimacy not only with the patient but also with the therapist's thinking about the patient. In addition, the verbatim comments of the supervisor appear on the same page in which the interview material is presented. Thus the book presents not only the interpretations and conclusions of the authors about a 40-year-old woman treated by brief, psychoanalytically oriented psychotherapy, but also includes a good deal of the raw data of treatment obtained not only from the therapist, but also from the patient and the supervisor. A final section of the book presents a verbatim account of several tests administered to the patient prior to therapy which were interpreted blindly and independently by two clinical psychologists. The authors' attacks upon the psychological reports are somewhat unconvincing in view of the limited testing and poor rapport with the patient during examination.

Apparently intended as something of a casebook for the earlier book by Dollard and Miller, *Personality and Psychotherapy* (1950), the introductory chapters of *Steps in Psychotherapy* summarize very briefly one learning theory interpretation of the therapeutic process and provide an extremely lucid résumé of the therapeutic principles presented in the earlier book. Although the 17 interviews are intended to illustrate the principles expounded, the relevance of the learning theory is not made particularly discernible in the supervisor's extended comments. The learning theory appears essentially to

be an apologia for the particular brand of brief analytic therapy which is presented. Only in the continued insistence on the therapist's duty to help the patient to "label" his experiences does there seem to be much extension of or departure from the usual procedures of brief, analytic treatment methods.

Regardless of one's point of view toward the brand of therapy advocated, the interested reader is in a position to inspect the supervisor's hypotheses and, to some extent, consider the evidence on which they were based as well as the effect the hypotheses had upon the patient when the therapist followed instructions. It is regrettable that space limitations prevented the authors from publishing more than selected and edited excerpts of the 16 interviews. Such questions as the following indicate the possibilities of using the book as a teaching device, although the answers can't be found in the book because of the limited material presented: (a) Was the early diagnosis of sexual maladjustment as the major problem justified, or did it stem from the theoretical bias of the supervisor? (b) Was the patient rewarded primarily for discussing sexual material and "punished" for other topics of discussion which might have been important? (c) Did the mother and husband play such minor roles in the patient's problems as the authors indicate?

In a charming and disarming sentence the authors "expect to find most helpful the views of those commentators who have themselves published their own uncensored remarks as therapists or supervisors." Although the drive revealed by this cue is commendable in a world where justice *should* reign supreme, the expectation is hardly cricket since it

reduces the universe of "qualified" commentators to perhaps five or six persons now living.

This reviewer will rest content with the observation that the edited case material fastens so insistently upon the several conflict theme that there is little opportunity for the reader to determine whether any alternative explanations of the patient's symptoms might have been tenable. Such a question does not, of course, detract from the real value of the book in illustrating a particular approach to therapy.

VICTOR C. RAIMY
University of Colorado

PICKFORD, R. W. *Individual differences in colour vision.* London: Routledge & Kegan Paul, Ltd., 1951. Pp. xviii + 386. 30s.

This book is primarily a detailed report of seven experiments on color vision carried out over a period of more than eight years with some 1100 observers. The research was designed to develop a diagnostic test of color vision, to help clarify the existing confused classifications of color blindness, and to test the adequacy of the Young-Helmholtz theory of color vision.

These aims were achieved with considerable success. Pickford's four-color test, which was used in his main experiment, turns out to be one which can be readily adapted to the clinical testing situation to give objective diagnostic information. His test data provide an operational basis for defining categories of normality, of color weakness, of anomalies, and of major defectives. His results point strongly toward a theory of color vision much like Houston's modification of the Hering theory, and can be fitted to the Young-Helmholtz theory only with unpar-

simonious additions to the basic trichomatic notion.

Although Pickford himself considers the work he has done as more preliminary than definitive, there are several important contributions which this research has made to the field of color vision testing. Primary among these is the fact that diagnostic categories are clear cut and objectively defined. A statistical procedure is used for defining "deviant" and "color weak" groups as falling respectively in the 1-to-2 and 2-to-3 sigma brackets of the normal distribution, and justifies separating them from the "anomalous" who deviate discontinuously out as far as the 5-to-7 sigma brackets. Protanopes and deuteranopes are distinguished in terms of their much greater color-matching variability and on a basis of their brightness matches. He reports that there are few "pure" dichromats (using the term in its traditional sense), but shows how the "nonpure" dichromats can be separated from the anomalous trichromats.

Another important contribution is the demonstration that individuals who may by some tests be considered normal, may at the same time show an independent weakness in any one of four spectral regions: red, yellow, green, and blue. Pickford shows, however, that there is likely to be some relationship between yellow and blue, or red and green sensitivity.

Of great practical value to those who must give color vision tests in industry or elsewhere are the precautions, described in detail, which show how the clinician should approach an individual to be tested so that the test will not be invalidated. The color defective has learned to use extra cues and all sorts of ruses to get by in his everyday life and frequently is a better examiner of the unwary tester than the tester is of him.

There will be some readers who will be critical of the general experimental technique used. From the standpoint of classical psychophysics, it may appear that the data are skimpy. Typically, one observation by a bracketing procedure was the only basis for scoring a single color-matching judgment, and a single match with each of four colors gave all the diagnostic information obtained for a single observer. There are, however, several factors which make these data considerably more than minimally acceptable. For one thing, the colorimetric technique used in the main experiment is inherently precise, and the bracketing procedure carried out by the experimenter was carefully designed to give a measure of variability as well as one of central tendency. One thing to remember is that the author intended to develop a practical testing procedure so that the whole test could be given in a matter of 10 or 15 minutes. Pickford did run a test-retest reliability series on 36 observers which gave coefficients ranging from .93 to .98 for the more important red and green test colors. The author intends to extend this elaborate and time-consuming preliminary study, and has already organized a survey of at least 20,000 British observers.

Although Pickford's classificational categories are objectively defined, they are specified in terms of the arbitrary scales of his colorimeter. It is unfortunate that he did not report internationally standardized psychophysical specifications (C.I.E.) for his test colors, and that he did not specify his data in these generally comprehensible units for comparative purposes.

Some criticism could be made of

the organization of the book. In general, a chronological sequence of experiments is followed, but it seems cluttered at times with repetition, with the inclusion of too much detail on the subjective reports of isolated individuals, and with introspective reports of the author.

In conclusion it may be said that no person who must test the color vision of others should omit studying this book carefully. As Professor Bartlett said in his Foreword to the book, "There is no other book I know in which the general methodology of this type of psychophysical investigation is dealt with anything like as well."

ROBERT W. BURNHAM
Eastman Kodak Company,
Rochester, New York

WERT, JAMES E., NEIDT, CHARLES O., & AHMANN, J. STANLEY. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954. Pp. viii+435. $5.00.

This is a lengthy but chatty introduction to descriptive and inferential statistics which progresses at a pedagogical pace suitable to the median student. Many detailed examples are well presented with emphasis on underlying rationale and assumptions. Arithmetic and high school algebra are the only requisites. The first five chapters present the customary procedures for collecting and summarizing data including central tendency, percentiles, variability, and correlation. Three chapters are then devoted to the concepts of sampling and

statistical inference. The remaining 11 chapters deal with the mean, analysis of variance and covariance, linear and curvilinear regression, and elaborations of correlation. Content-wise the book's chief claim to attention in the elementary market place is its more than ample discussion, and certain unusual material such as double classification analysis of variance with disproportionate frequencies and the application of the two-class discriminant function. A worthy contender for the first two or three semesters.

LEONARD S. KOGAN
Institute of Welfare Research,
Community Service Society of
New York

UNDERWOOD, BENTON J., DUNCAN, CARL P., TAYLOR, JANET A., & COTTON, JOHN W. *Elementary statistics.* New York: Appleton-Century-Crofts, 1954. Pp. ix + 239. $3.25.

This little book is designed for use in the introductory undergraduate one-semester course. Coverage is classic and includes handling of distributions, percentiles, central tendency, variability, correlation, the normal curve, significance of differences between means, simple analysis of variance, and chi square. It is well written and pitched toward problems of inference in research as are a number of other elementary texts.

LEONARD S. KOGAN
Institute of Welfare Research,
Community Service Society of
New York

ADAMS, DONALD K., *et al.* *Learning theory, personality theory, and clinical research.* (The Kentucky Symposium.) New York: Wiley, 1954. Pp. ix+164. $3.50.

ARNHEIM, RUDOLF. *Art and visual perception: a psychology of the creative eye.* Berkeley: Univer. of California Press, 1954. Pp. x+408. $10.00.

ARSENIAN, SETH, & MCKENZIE, FRANCIS W. *Counseling in the YMCA.* New York: Association Press, 1954. Pp. 126. $2.00.

BARRELL, JOSEPH. *A philosophical study of the human mind.* New York: Philosophical Library, 1954. Pp. xii+575. $6.00.

BECK, SAMUEL J. *The six schizophrenias: reaction patterns in children and adults.* Research Monograph No. 6, American Orthopsychiatric Association. New York: American Orthopsychiatric Association, 1954. Pp. viii+238. $5.00.

BISCHOF, L. J. *Intelligence: statistical concepts of its nature.* Garden City, N.Y.: Doubleday, 1954. Pp. vi+33. $.85.

BRAATØY, TRYGVE. *Fundamentals of psychoanalytic technique.* New York: Wiley, 1954. Pp. xi+404. $6.00.

BROWN, J. A. C. *The social psychology of industry.* Baltimore: Pelican, 1954. Pp. 309. $.65.

BURT, SIR CYRIL. *The causes and treatment of backwardness.* New York: Philosophical Library, 1953. Pp. 128. $3.75.

CAMPBELL, PAUL, & HOWARD, PETER. *Remaking men.* New York: Arrowhead Books, 1954. Pp. 126. $1.50.

DRIVEE, HELEN IRENE. *Multiple counseling . . . a small-group discussion method for personal growth.* Madison, Wis.: Monoma Publications, 1954. Pp. 280. $5.00.

ENGLISH, O. S., & FINCH, S. M. *Introduction to psychiatry.* New York: W. W. Norton, 1954. Pp. viii+621. $7.00.

ESTES, WILLIAM K., *et al.* *Modern learning theory.* New York: Appleton-Century-Crofts, 1954. Pp. xi+379. $5.00.

FRANK, LAWRENCE K., HARRISON, ROSS, HELLERSBERG, ELISABETH, MACHOVER, KAREN, & STEINER, META. *Personality development in adolescent girls.* Monograph, Society for Research in Child Development, Vol. 16, No. 53, 1951. Yellow Springs, Ohio: Antioch Press, 1953. Pp. 316.

FRIEDMANN, E. A., HAVIGHURST, R. J., *et al.* *The meaning of work and retirement.* Chicago: Univer. of Chicago Press, 1954. Pp. vii+197. $3.75.

FULLER, JOHN L. *Nature and nurture: a modern synthesis.* Garden City, N. Y.: Doubleday, 1954. Pp. vi+40. $.85.

GOOD, CARTER V., & SCATES, DOUGLAS E. *Methods of research: educational, psychological, sociological.* New York: Appleton-Century-Crofts, 1954. Pp. xx+920. $6.00.

GUTKIND, E. A. *Community and environment: a discourse on social ecology.* New York: Philosophical Library, 1953. Pp. xiii+81. $3.75.

HARPER, ARTHUR EDWIN, JR. *Differential patterns in schizophrenia on the Wechsler-Bellevue intelligence test.* Allahabad, U. P., India: Allahabad Christian Press, 1950. Pp. 71.

HIMES, JOSEPH S. *Social planning in America.* New York: Doubleday, 1954. Pp. ix+59. $.95.

HOCH, PAUL H., & ZUBIN, JOSEPH. (Eds.) *Depression.* New York: Grune & Stratton, 1954. Pp. x+277. $5.50.

HODGSON, KENNETH W. *The deaf and their problems.* New York: Philosophical Library, 1953. Pp. xx+364. $6.00.

HOFSTATTER, PETER R. *Einführung in die Sozialpsychologie.* Humboldt-Verlag, 1954. Pp. 535. DM 12.50.

HUMPHREYS, J. ANTHONY, & TRAXLER, ARTHUR E. *Guidance services.* Chicago: Science Research Associates, 1954. Pp. xvii+438. $4.75.

ITTELSON, W. H., & CANTRIL, H. *Perception: a transitional approach.* Garden City, N.Y.: Doubleday, 1954. Pp. ix+33. $.85.

LAYTON, WILBUR L. (Ed.) *Selection and counseling of students in engineering.* Minnesota Studies in Student Personnel Work, No. 4. Minneapolis: Univer. of Minnesota Press, 1954. Pp. 89. $1.75.

LAZARSFELD, PAUL F. (Ed.) *Mathematical thinking in the social sciences.* Glencoe, Ill.: The Free Press, 1954. $10.00.

LEUBA, CLARENCE. *The sexual nature of man and its management.* Garden City, N. Y.: Doubleday, 1954. Pp. vii+40. $.85.

MCMILLAN, BROCKWAY, GRANT, DAVID A., FITTS, PAUL M., FRICK, FREDERICK C., MCCULLOGH, WARREN S., MILLER, GEORGE A., BROSIN, HENRY, W. *Current trends in information theory.* Pittsburgh: Univer. of Pittsburgh Press, 1953. Pp. xi+188. $4.00.

ODLUM, DORIS M. *Psychology, the nurse and the patient.* (2nd Ed.) New York: Philosophical Library, 1954. Pp. 168. $4.75.

PEARSON, GERALD H. J. *Psychoanalysis and the education of the child.* New York: W. W. Norton, 1954. Pp. x+357. $5.00.

PIAGET, JEAN. *The construction of reality in the child.* (Trans. by Margaret Cook.) New York: Basic Books, 1954. Pp. xiii+386. $6.00.

RIESMAN, DAVID. *Individualism reconsidered.* Glencoe, Ill.: The Free Press, 1954. Pp. 529. $6.00.

ROSS, C. C. *Measurement in today's schools.* (3rd Ed.) (Rev. by Julian C. Stanley.) New York: Prentice-Hall, 1954. Pp. xv+485. $5.00.

RUSSELL, DAVID H. *The dimensions of children's meaning vocabularies in grades four through twelve.* Univer. of Calif. Publ. in Education, Vol. 11, No. 5. Berkeley: Univer. of California Press, 1954. Pp. 315–415. $1.25.

SARASON, SEYMOUR B. *The clinical interaction.* New York: Harper, 1954. Pp. x+425. $5.00.

SAVAGE, LEONARD J. *The foundations of statistics.* New York: Wiley, 1954. Pp. xv+294. $6.00.

SENN, MILTON J. E. (Ed.) *Problems of infancy and childhood.* Transactions of the Seventh Conference, March 23 and 24, 1953, New York, N. Y., Josiah Macy, Jr. Foundation. New York: Josiah Macy, Jr. Foundation, 1954. Pp. 196. $2.75.

SMITH, HENRY P. *Psychology in teaching.* New York: Prentice-Hall, 1954. Pp. xiii+466. $4.95.

SNYDER, RICHARD C., BRUCK, H. W., & SAPIN, BURTON. *Decision-making as an approach to the study of international politics.* Princeton: Princeton Univer. Press, 1954. Pp. 120.

SOLOMON, JOSEPH C. *A synthesis of human behavior.* New York: Grune & Stratton, 1954. Pp. xii+265. $5.50.

STERNEGGER, BENEDIKT. *Das verlorene Wort.* Augsburg: Hans Rosler, 1954. Pp. 84.

SULLIVAN, HARRY S. *The psychiatric interview.* (Helen Swick Perry and Mary Ladd Gawel, Eds.) New York: W. W. Norton, 1954. Pp. xxiii+246. $4.50.

THELEN, HERBERT A. *Dynamics of groups at work.* Chicago: Univer. of Chicago Press, 1954. Pp. ix +379. $6.00.

TINBERGEN, NIKO. *The herring gull's world.* New York: Frederick A. Praeger, 1953. Pp. xvi+255. $4.00.

UNDERWOOD, BENTON J., et al. *Elementary statistics.* New York: Appleton-Century-Crofts, 1954. Pp. ix+239. $3.25.

WARTERS, JANE. *Techniques of counseling.* New York: McGraw-Hill, 1954. Pp. viii+384. $4.75.

WATSON, ROBERT I. *Psychology as a profession.* Garden City, N. Y.: Doubleday, 1954. Pp. x+65. $.95.

WILLIAMS, J. D. *The compleat strategyst; being a primer on the theory of games of strategy.* New York: McGraw-Hill, 1954. Pp. xiii+234. $4.75.

WOLSTEIN, BENJAMIN. *Transference; its meaning and function in psychoanalytic therapy.* New York: Grune & Stratton, 1954. Pp. xiii+206. $5.00.

WOODWORTH, ROBERT S., & SCHLOSBERG, HAROLD. *Experimental psychology.* (Rev. Ed.) New York: Holt, 1954. Pp. xi+948. $8.95.

ZUBEK, JOHN P., & SOLBERG, PATRICIA A. *Human development.* New York: McGraw-Hill, 1954. Pp. vii+476. $6.00.

INDEX OF AUTHORS

ARTICLES

BOOKS REVIEWED

BOOK REVIEWERS